DOCUMENT RESUME

ED 082 797 LI 004 519

AUTHOR	Mauerhoff, Georg R., Comp.
TITLE	Open Conference on Information Science in Canada, Proceedings (1st, Montebello, Quebec, May 14 & 15, 1973).
INSTITUTION	Canadian Association for Information Science, Ottawa (Ontario).
PUB DATE	73
NOTE	186p.; (100 references)
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Communications; Conference Reports; Education; Foreign Countries; *Information Networks; *Information Science; Information Scientists; *Information Systems; Library Education; Policy; Research; Technology
IDENTIFIERS	*Canada; Information Science Education

ABSTRACT

The papers presented at the first Canadian conference on information science are presented in this volume. Six presentations were given under the general topic of research: Research into Privacy and Data Banks, Communications Knowledge Software Industry for Canada, Census Data Access and Statistical Information Management, Communication System Habitability: The Need for Behavioural Research, FABS (Formulated Abstracting): An Experiment in Regularized Content Description, and I.F. Thesaurus of Building Science and Technology. The subject of education included: Education for Information Science: An Apologia and an Introduction, National Goals for the Education and Training of Personnel for a National System for Scientific and Technological Information, Data--The Missing Link in the Evolution of Library Science Curricula, Teacher Education in Information Systems Utilization, and Laval University's Experience with Data Bases. Papers on technology are: Computer Communications in Canada, Communications and the Wired City, Modularity - Microform and Microcomputers, A Hybrid Interactive Search System, Optical Character Recognition in Information Processing, A UDC Water Thesaurus/Concordance: Development and Use, and An Urban and Regional Information System, as Applied to the Eastern Townships Region, Province of Quebec. Four papers are also included on policy and information networks in Canada. (SJ)

ED 082797

L'ASSOCIATION CANADIENNE DES SCIENCES DE L'INFORMATION

CANADIAN ASSOCIATION FOR INFORMATION SCIENCE

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

COMPTE RENDU

PROCEEDINGS

compilé par — compiled by

Georg R. Mauerhoff

PREMIÈRE CONFÉRENCE PUBLIQUE SUR LES SCIENCES DE L'INFORMATION AU CANADA

FIRST OPEN CONFERENCE ON INFORMATION SCIENCE IN CANADA

MONTEBELLO, QUÉBEC

LES 14 ET 15 MAI 1973

MAY 14 & 15, 1973

LI 004 519

L'Energie atomique du Canada Limitée

Atomic Energy of Canada Limited

Chalk River, Ontario
1973

FILMED FROM BEST AVAILABLE COPY

PRÉSIDENT DE LA CONFÉRENCE:	F.T. Dolan
VICE-PRÉSIDENT:	P. Wolters
COMITÉ DU PROGRAMME:	Georg Mauerhoff (président) G.X. Amey C.D. Batty S. Heaps J.E. Brown

CONFERENCE CHAIRMAN:	F.T. Dolan
VICE-CHAIRMAN	P. Wolters
PROGRAMME COMMITTEE:	Georg R. Mauerhoff (Chairman) G.X. Amey C.D. Batty S. Heaps J.E. Brown

PROGRAMME

Sunday, May 13

8:30 p.m. - 10:30 p.m.	-MIXER
2:00 p.m. -	-REGISTRATION

Monday, May 14

9:00 a.m. - 12:30 p.m.	-OPENING REMARKS -SESSION I - RESEARCH
12:30 p.m. - 2:00 p.m.	-LUNCH -MR. STEPHEN FRANKLIN "Knowledge Park"
2:00 p.m. - 5:00 p.m.	-SESSION II - EDUCATION
5:30 p.m. - 6:30 p.m.	-COCKTAILS
8:00 p.m. - 9:00 p.m.	-BUSINESS MEETING

Tuesday, May 15

9:00 a.m. - 12:30 p.m.	-SESSION III - TECHNOLOGY
12:30 p.m. - 2:00 p.m.	-SMORGASBORD
2:00 p.m. - 4:00 p.m.	-SESSION IV - POLICY & NETWORKING -CLOSING REMARKS

P R O G R A M M E

Le dimanche 13 mai

14h -	-INSCRIPTION
20h30 - 22h30	-RÉUNION INTIME

Le lundi 14 mai

9h - 12h30	-PAROLES D'OUVERTURE -Ie SÉANCE - RECHERCHE
12h30 - 14h	-DÉJEUNER -M. STEPHEN FRANKLIN "Knowledge Park"
14h - 17h	-IIe SÉANCE - ÉDUCATION
17h30 - 18h30	-COCKTAIL
20h - 21h	-RÉUNION D'AFFAIRES

Le mardi 15 mai

9h - 12h30	-IIIe SÉANCE - TECHNOLOGIE
12h30 - 2h	-BUFFET
2h - 4h	-IVe SÉANCE - DIRECTIVES ET RÉSEAUX DOCUMENTAIRES -PAROLES DE CLÔTURE

TABLE DES MATIERES

TABLE OF CONTENTS

PROGRAMME — Page

RESEARCH/RECHERCHE

Introduction

Research into Privacy and Data Banks - A. Gotlieb 1

Communications Knowledge Software Industry for Canada - G.B. Thompson . 2

Census Data Access and Statistical Information Management (Accès aux données du recensement et gestion de l'information statistique) - J.G. Stinson 9

Communication System Habitability: The Need for Behavioural Research (L'habitabilité des systèmes de communication: le besoin d'une recherche du comportement) - A. Pietrzyk 10

FABS (Formulated Abstracting): An Experiment in Regularized Content Description (FABS: une étude pour la formalisation du résumé signalétique) - B. Harris and T.R. Hofmann 17

I.F. Thesaurus of Building Science and Technology (Thésaurus I.F. - sciences et technologie du bâtiment) - C.H. Davidson 28

EDUCATION

Education for Information Science: An Apologia and an Introduction (L'éducation pour les sciences de l'information: apologia et une introduction) - C.D. Batty 43

National Goals for the Education and Training of Personnel for a National System for Scientific and Technological Information (Objectifs nationaux de la formation de personnel en vue d'un système national d'information scientifique et technique) - D.H. Evans 47

Data - The Missing Link in the Evolution of Library Science Curricula (Les donnés - l'anneau manquant dans l'évolution des programmes d'études en bibliothéconomie) - F.T. Dolan 54

Teacher Education in Information Systems Utilization (Préparation des professeurs pour l'utilization des systèmes d'information) - H.J.A. Goodman 65

L'expérience de l'Université Laval dans l'utilisation des banques d'information (Laval University's Experience with Data Bases) - R. Francoeur, E. Loranger and A. Gamache 72

TECHNOLOGY/TECHNOLOGIE

Introduction

Computer Communications in Canada - D. Horton

Communications and the Wired City - L. Durr 81

Modularity, Microform & Microcomputers - J.W. Cherry and E.V. Goodman . 82

Page

A Hybrid Interactive Search System (Un système hybride interdépendant de la recherche documentaire) - J.D. Currie 90

Optical Character Recognition in Information Processing (Reconnaissance optique des caractères imprimés informatique résumés) - C.Y. Suen 97

A UDC Water Thesaurus/Concordance: Development and Use (La concordance de CDU avec un thésaurus de l'eau: développement et utilisation) - H.E. McCuaig and M.A. Mercier 104

Un système d'information urbain et régional, appliqué à la région de l'estrie, province de Québec (An Urban and Regional Information System, as Applied to the Eastern Townships Region, Province of Quebec) - P. Lacasse 119

POLICY AND INFORMATION NETWORKS/DIRECTIVES ET RÉSEAUX DOCUMENTAIRES

Introduction - J.E. Brown 134

Science Policy and Scientific Information in Canada - A. Beaulnes

Evolution of Canadian System for Geoscience Data, 1964-73 (Evolution du système canadien de données géoscientifiques, 1964-73) - C.F. Burk, Jr. 138

The Potential of Regional Government for the Rationalization of Local Information Systems (Le potentiel offert par un gouvernement régional dans la rationalisation des systèmes d'informations locaux) - J.N. Brown 148

Le Système Canadien de Renseignements sur la Construction (A Construction Information System for Canada) - O. Dupuis 161

APPENDIX/ANNEXE

Other contributed papers /Autres articles soumis 168

RESEARCH INTO PRIVACY AND DATA BANKS

A.E. Gotlieb
Deputy Minister
Department of Communications
Ottawa, Ontario K1A 0C8

ABSTRACT

The issue of privacy represents, in many respects, a test issue or case study of social and cultural problems in a post-industrial, information-oriented, society. The issue has been created, in part, as a direct consequence of the efficiency of computers as information processors, and as an unintended product of the social benefits generated by computers; it has been created in part also as an expression of public concern about the relationship between computer power and the potentially manipulative power of institutions operating information systems. The issue has been the subject of extensive study in all advanced, western countries. The Canadian Departments of Communications and Justice jointly sponsored a Task Force inquiry, of which the report was published in December, 1972. A special inter-departmental committee is at present studying rules to protect privacy which could be applied to data banks operated by the federal government. Consideration is being given also to parallel issues such as the extra-territorial location of data banks.

COMMUNICATIONS KNOWLEDGE SOFTWARE INDUSTRY FOR CANADA.

Gordon B. Thompson
Bell Northern Research
Ottawa, Canada.

ABSTRACT

Canada's unique advantages and problems likely to be encountered in the creation of a communications knowledge software industry are examined in the light of two recent documents, one Japanese and one Canadian, on the topic of computers and communications.

Innovation, according to one of our Canadian idea generators breaks down into two classes with respect to the treatment of labour. Eric Burke of the University of Waterloo states that insufficient attention has been given to the labour absorbing kind of innovation to balance the amount of labour releasing innovation we are practising. In describing these two classifications of innovation, he has drawn our attention to the relationship between the innovative imbalance that exists in our society in this regard, and economic instability. There is no real risk in an innovation that will release labour, as in a cost reduction innovation. On the other hand, labour absorbing kinds of innovation are generally more complex, and contain more uncertainty and risk. Canadian conservatism favours the patch-up labour releasing kind of innovation. We are really very good at it. Now, according to Burke, we must get very much better at the more complex kind of innovation if we want a healthy economy.

What sort of labour absorbing kinds of innovation can be created and brought into wide application that don't require complex organizational structures for each individual innovation? Does such a class of innovative activities exist? Our studies indicate that it indeed does, and miniscule examples of such activities already exist to-day. The Quaternary or communications knowledge software industry is the area we have been investigating.

Some writers have described this Quaternary sector as the knowledge sector, but perhaps this is only a temporary name, for it doesn't catch the essence of the unique characteristic of this fourth sector. This sector,

which really can't exist without the first three being very well developed somewhere in the world, can in itself produce wealth. It is the wealth creating aspect that is both relatively new and potentially very significant.

The Japanese view the knowledge industry as becoming the most significant sector of their economy during the next decade. Ando, (1972). They define this knowledge industry as a portion of both the secondary and tertiary economic sectors of old. Those hardware products that relate to the knowledge industry, like computers for example, are included, along with education, broadcasting and publishing. Hence, their definition of the knowledge industry includes both hardware and software products. As a result the unique characteristics of the software portion of this industry will remain hidden from their view for some time. They have rightly identified the knowledge industry as both big and very significant, however they may have partitioned or defined the sector incorrectly. In so doing, the real wealth producing characteristic remains less well monitored and understood than if the partitioning had been otherwise.

In their model of the world's transactions, the Club of Rome researchers didn't include the possibility of a significant and increasing portion of the world's transactions being both non-polluting and capital or wealth producing. For small levels of this activity, the "World Two" model, reported by Meadows (1972), is quite adequate. But, should we learn to make this activity really run, then the future looks very much brighter than the authors have forecast.

Software transactions already exist to-day. They are not really new. Everytime a song is played on a radio or television station, there is a transaction, and money flows through a network to the author and composer. Galt MacDermot worked for two weeks to produce the musical score for "Hair", and his ownership of the rights, which is a sort of virtual capital, has earned him over a million dollars. He can go to a political convention, while his rights continue to earn his keep. He is one of the new kind of capitalists. The challenge is to create an environment that stimulates more and more people to discover more and more ways of doing the same kind of thing.

By building a transaction sensitive network embracing our communications and computer technology skills, an environment can be created that would allow one to produce software products and offer them for sale by investing principally time and effort. One wouldn't need to set up a corporation to manufacture, distribute and market one's

creation as is the case of innovative activity in the secondary sector. The manufacturing, distribution and marketing functions in the Quaternary sector are shared very effectively.

Canada is in a unique position to lead the world in this new economic activity. First of all, we are world leaders in terms of per capita installed communications facilities. Secondly, we are world leaders in understanding something about the significance and utility of the communications and computer technologies. We probably have a decade of very good times ahead of us to provide the funds necessary to embark upon research and implementation programs. With all of this going for us, maybe we even have a responsibility to search out the potential of this Quaternary sector of economic activity.

Illich (1972) suggests that communications networks should serve society in a manner that aids in the discovery of mutual interests amongst the users. Robert Theobald holds similar opinions of the value of dynamic and responsive networks. This is one of the properties that could be designed into the communications knowledge software industry should we so desire. Both writers claim rather fantastic payoffs for such efforts, and our analysis would tend to support their views. However, we have one or two experimental situations that point out the extreme difficulty in actually creating a network that gets used the way these authors describe. There is a need for some quite sophisticated research here. The issues are not clear, and the evaluation parameters are just non-existent. However, continued development of the simplistic broadcast syndrome seems both unimaginative and counterproductive. It is expected that work in this area in the next year will cause a revision of Thompson's (1970) third characterization of communications innovations.

Why insn't there more research in these areas? Why don't the Department of Communications outputs discuss these larger issues? It is my hypothesis that the structure of the organization within which the inquiring group is embedded, largely determines the kind of output that will be produced. Our increasing dependence upon prestigous research done in institutionalized centers may be an even greater inhibitor of the communications knowledge industry than our Canadian tradition of conservatism in innovation.

When Marshall McLuhan says the medium is the message, he is referring to a characteristic of mass media. However, this principle may well be more general. It seems also to apply to the various ways we study and report within our

society.

Royal commissions, government task forces, and other similar bodies seem constrained to produce output that is free of significant surprises. Perhaps this somewhat conservative characteristic is due to the environment in which the inquiring organization finds itself, and so is determined at the inception of the inquiry.

As a case in point, a comparison, albeit biased, of the output of Canada's Task Force on Computer Communications with that of the Japan Computer Usage Development Institute shows our Canadian document as a rather mild mannered innocuous statement, while the Japanese one is a strong statement relating to a real hope of survival for the Japanese over the next quarter century. The comparison is rather like that between lemonade and whisky. But how can you make whisky in a juice squeezer? The message from the juice squeezer can only be lemonade, never whisky! In this sense, the medium is the message.

The Japanese report suggests that as a national goal of the highest priority, it is mandatory to change their way of thinking, from one based on "the industrial mind" to one based on "the computer mind". They are not unaware of the extent of the change this involves. They state that in order to meet their objectives, they will require a cumulative investment of $65 billion by the year 1985, and that this amount of capital can be drained off the economy if they maintain a yearly GNP growth of eleven percent. One measure of the changes they expect to produce is that by 1985 they hope to see fifteen percent of the average wage-earner's expenditures devoted to services he receives from the computer communications industry. In short, they relate post-industrial survival to the successful establishment of an information based economy.

The Canadian Task Force on Computer Communication had no mandate to go so far. A new business seemed to be opening, so a government sponsored organization was established to do a conventional analysis of this business. They did an admirable job. Their report is really very good, given the restrictions placed on the alternatives they might have chosen. They were part of an organization whose rate of causing change is, by definition, geared to what the majority of the Canadian population will accept. This rate of change can hardly be described as radical.

It is probably very safe to say that predictions about computers and communications are either very much overstated or else very much understated. There is no

significant probability that our perceptions can be right on target. Given the bias to conventional extrapolative forecasting that pervades the organizational structure in which the task force was embedded, and the duality of the future they are dealing with, little question remains as to why the output was "so Canadian".

But, isn't this what government is for? Should not society's minimum acceptable goals of advancement come from its democraticlly chosen government? The risky and uncertain advances must come from organizations that can be killed if they are wrong. Killing a government is dangerous and expensive. Killing a corporation is called bankruptcy, and can even be profitable. Hence, a study done under government sponsorship can hardly be expected to be dangerous.

On the other hand, large corporate interests don't have the public trust nor the capital needed to get truly new, large and complex ideas launched, particularly if these ideas involve public attitude changes.

Universities really can't help here, for their role relates to teaching, identification and preservation of that that is accepted as truly valuable. They are in a poor position to embrace risky enterprise. Too much is at stake if they embrace an idea that turns out to be foolish.

If government task forces are inhibited from producing risk laden, imaginative and fundamentally new outputs, corporations can't because of capital and believability constraints, and Universities find the job too risky, who can do this job competently in our society?

There remain the think-tank institutes that have been proposed, with the Ritchie Report being an example of such a proposal. These institutes are likely to produce no real force for change, and like the Hudson Institute, become concerned with their own survival to an extent that really affects their ability to act in a dangerous or preposterous way. Any prestigous organization is vitally concerned with the preservation of that prestige. The right to fail gloriously must be one of the characteristics built into any organization that is to do a really effective job at producing predictive output that relates to basic changes that might significantly alter our present course of development. Perhaps we need some new structures for this kind of research.

Returning to the comparison between the Canadian and Japanese computer documents, one can observe that the

difference is most probably due to the totally different organizational environment in which the two groups were embedded. The Japanese situation was that of a blatantly biased, special purpose, multi-sponsored, institute of limited life. The Canadian group found itself bound into a more conventional situation, and so their output, perforce, represents a reasonable minimum level of expectation. On the other hand, the Japanese document, because of their particular organizational arrangement, presents something approaching their view of an upper bound.

The information marketplace business is like an atomic reaction. It has a critical mass that must be exceeded before it can carry on by itself. In the upper bound studies we have made, the critical size that is required before economic self-sufficiency is reached seems very large. The scale on which the thinking must be done is uncomfortably large. This is perhaps why the wealth generating characteristics of a purely software kind of marketplace have gone unheralded for so long. It is possible that even the Japanese upper bound view can turn out to be too small.

Clearly there are many problems to be overcome before Canadians can claim to have established a knowledge software industry of significance. Few countries have the advantages we have, and none on a per capita basis can come close. There is much research to be done, research that seems to be less context free than those of us with a scientific inclination would prefer. The challenge is great, but the scenario at the end looks so rewarding.

REFERENCES.

ANDO, Kaora; Speech, "Social and Economic Implications of Computers and Communications"; International Conference on Computer Communication, October 1972, Washington, D.C.

"Branching Out", Report of the Canadian Computer/Communications Task Force. Information Canada, 1972.

ILLICH, Ivan; "Deschooling Society", Page 28, Harper and Row. 1972.

MEADOWS, D.H. etc.; "The Limits To Growth", Universe Books, New York, 1972.

"Plan for Information Society, The", Japan Computer Usage Development Institute, Tokyo, 1972.

THOMPSON, G.B.; "Moloch or Aquarius", THE, Issue 4, November 1970, Bell Northern Research.

CENSUS DATA ACCESS AND STATISTICAL INFORMATION MANAGEMENT (ACCES AUX DONNEES DU RECENSEMENT ET GESTION DE L'INFORMATION STATISTIQUE)

J.G. Stinson
Census Branch Statistics Canada
Ottawa, Ontario K1A OT7

ABSTRACT

The Census Branch at Statistics Canada has undertaken an expanded programme of tabulations from the 1971 Census. The size of this programme has necessitated the development of automated systems to produce directories of the data produced and the available formats. A major consideration in the planning and implementation of the programme and distribution of the data was the application of a different type of information management to ensure that the confidentiality requirements of the Statistics Act are maintained. (La Direction du recensement de Statistique Canada a institué un programme élargi de totalisations fondées sur les données du recensement de 1971. L'importance de ce programme a nécessité la mise au point de systèmes automatisés pour la production de répertoires des données produites et des modes de présentation disponibles. L'application d'un type différent de gestion de l'information a compté parmi les grandes considérations dans la planification et la mise en oeuvre du programme et dans la distribution des données, compte tenu de la nécessité de respecter le secret statistique.)

COMMUNICATION SYSTEM HABITABILITY: THE NEED FOR BEHAVIOURAL RESEARCH (L'HABITABILITÉ DES SYSTÈMES DE COMMUNICATION: LE BESOIN D'UNE RECHERCHE DU COMPORTEMENT)

Alfred Pietrzyk
Center for Applied Linguistics
Arlington, Virginia 22209

ABSTRACT

Modern systems for the communication of information may satisfy or frustrate their users depending on the habitability of given system environments. Communication system habitability, which entails many facets of individual and social user behaviour, is not being explored in a coherent framework. Studies of user behaviour in such environments as telecommunications call for nontrivial behavioural criteria. This paper is focused on Canadian needs for pragmatic research strategies and project inventories--first steps toward a refinement of behavioural criteria. (Les systèmes modernes pour les communications d'information peuvent satisfaire ou frustrer leurs usagers selon l'habitabilité de l'environnement du système donné. Cette habitabilité qui comprend plusieurs aspects du comportement individuel et social de l'usager n'est pas étudiée dans un cadre d'activité homogène. Les études du comportement de l'usage dans de telles conditions comme dans la télécommunication, exige des critères surs du comportement. Ce document pointe les besoins canadiens pour une stratégie d'application pratique de recherche et de relevé de projets--les premiers pas vers un raffinement des critères du comportement.)

In 1960, Mooers argued that customer use of an information system tends to be inversely proportionate to the amount of "pain" which access to information entails. We can utilize such essentially behavioural arguments to shed light on certain research priorities. In the following sketch, I shall attempt to highlight the need for behavioural research focused on uses of information systems in Canada. I shall include some comments on workable preliminaries.

CONTEXT

Information systems are considered to be systems for the communication of information. The term "communication" is, in this context, used in the classical sense entailing a flow of messages from informa-

tion producers to information consumers through appropriate channels.

My remarks are confined to scientific and technical information. This is not meant to imply that Canadian users of such information consist entirely of professional researchers and practitioners. Information scientists are becoming increasingly sensitive to the needs of non-professional audiences--a trend which is especially evident in recent publications by Heaps and Ingram (1971, 1972), Ironside (1971), and several others.

We can, by now, take it for granted that individual and social user behaviour--the human aspect--is crucial to the intelligent design and long-range maintenance of a modern communication system. In the Telecommission (1971) Report, the "social and environmental" aspects of new communication technologies are included among special priorities which call for "domestic innovative capacity." In terms of human consequences, the scenarios range from an Orwellian disregard for privacy to a utopian abundance of choices. Most of us look forward to communication innovations whose human impact is in some sense likely to be beneficial. However, we find it difficult to agree on specific kinds of desirable benefits.

In my view, progress in this complex area depends on the extent to which our collective wisdom can be based on behavioural research within a coherent framework. I shall therefore address myself briefly to the need for conceptual integration and, thereafter, consider the environmental viewpoint and the concept of habitability as possible catalysts toward conceptual unification.

NEED FOR CONCEPTUAL INTEGRATION

Behavioural research in scientific and technical communication is fragmented: it tends to cluster around a number of different conceptual models entailing different theoretical and methodological viewpoints. This conceptual fragmentation is clearly exemplified in the Annual Review of Information Science and Technology (Cuadra 1966-) which has dealt with behavioural matters in its regularly recurring chapter on "Information Needs and Uses" and to some extent also under such occasionally featured headings as "Man-Computer Communication," "Automated Language Processing," and "Design and Evaluation of Information Systems."

The conceptual organization of the "Needs and Uses" chapter has to a significant degree followed the outline introduced by Paisley (1968). The scientist is considered within nine communication-behaviour settings: his own head, his work team, a formal organization, an invisible college, his reference group, a membership group, a formal information system, a political system, and his culture. This kind of paradigm is, of course, not meant to account for the deeper complexities of scientific

and technical communication behaviour. It is merely a taxonomic summary which tells us nothing about the dynamic interplay among the various behavioural aspects.

The variables involved in scientific and technical communication behaviour are complex, numerous, and for the most part ill-defined. If the concepts and criteria of a narrow behavioural framework are applied to system design, we may have to face the consequences of suboptimization: certain user requirements within, for example, Paisley's (ibid.) nine behavioural settings may not be satisfied. The need for a comprehensive and coherent behavioural framework is obvious from, among other things, the CAN/SDI user study findings of Wolters and Brown (1971) and Mauerhoff (1971).

THE ENVIRONMENTAL VIEWPOINT

It may not be reasonable to expect early proposals for a specific unified framework. We can, nevertheless, speculate about the broader orientations. Among the more suitable options, the environmental viewpoint appears to be especially promising. I am not suggesting this merely because it is fashionable to think in these terms. I hope to show that the environmental emphasis may point to a working concept of considerable cohesion and operational scope.

Concerns referred to as environmental are discussed in many contexts in the Telecommission (1971) Report. Environmental interests are being pursued deliberately in some government programmes dealing with communication. These are significant trends which, however, seem to lack an explicit emphasis on one crucial aspect: the systematic nature of communication environments. This is understandable in light of the fact that an ecology of communication accounting for the systematicity of communication environments does not yet exist in any meaningful sense.

A theoretically sophisticated ecology of communication might some day be added to the burgeoning pursuits of human ecology and man-environment studies. In such a context, the behavioural value systems to be explored in an ecology of communication might eventually be related to the quality of life in general--something that would have a liberating effect on our current parochially oriented "needs and uses" research. From a theoretical standpoint, the explanatory potentials of an ecology of communication may ultimately depend on concepts that are in some way analogous to those used in general ecology; the communication system may come to be viewed in terms of a dynamic ecosystem (Whittaker 1971).

Habitability

In order to account for the impact of the environment on organisms, ecologists use the holocoenotic principle. Holocoenosis pertains to those environmental factors which act as a complex whole rather than

separately and independently (Platt 1971). Using this principle as a general guide, we can outline a working concept--habitability--for the integrated study of scientific and technical communication behaviour.

In my suggestive definition, the habitability of a given communication system is equivalent to the aggregate or profile of all evaluations which characterize the environmental impact of this system on user behaviour. Thus, among candidates for inclusion in the composite assessment referred to as habitability are all qualitative and quantitative evaluations in terms of user satisfaction, frustration, cognitive responses, convenience, privacy, access, overload, research duplication, productivity, savings, etc. The evaluations would include user perceptions of environmental attributes and system performance in terms of quality, relevance, recall, precision, speed, novelty, coverage, selectivity, language, time, pricing and many other still poorly understood variables (cf. King 1968). It is, moreover, assumed that such evaluations are, in important ways, relatable to one another and to habitability as a whole. This would reduce the implications for suboptimization mentioned earlier. The term "habitability" is not unprecedented in the information science literature; its previous definition and use by Watt (1968) were, however, confined to the special case of man-machine interface languages.

My conception of habitability has tangible practical implications for the analysis of user behaviour. The initial tool for habitability assessment would be relatively simple as far as its general parameters are concerned: it would take the form of a two-dimensional environmental impact matrix of the kind employed in regional and industrial development surveys (e.g. Truswell 1972). One of the two basic parameters would consist of factors in a particular communication environment. The taxonomy of these factors may range from macroanalytic categories, such as the generic settings of Paisley (1968) referred to earlier, to the microanalytic performance details of given system functions and components such as the search protocols, keyboard features, and display capabilities of telecommunications terminals. The second parameter would specify the users' behavioural variables--satisfaction, frustration, etc.--within appropriate sets. The grid spaces of a habitability matrix would contain qualitative and quantitative evaluations of the impact of given environmental factors on given behavioural variables. The resulting habitability profile could be further analyzed for distinctive features in its overall gestalt. Because of the simplicity at their general level, habitability matrices would be conceptually open-ended and hospitable to diverse research trends. They would provide for variable foci and degrees of conceptual detail. Eventually, computer processing may be needed in order to deal with the large number and complexity of items in certain matrices.

Communication system habitability is a relative concept comparable

to that proposed for the habitability of dwellings (Fraser 1969). This means that the habitability of given communication innovations in Canada may vary in significant ways depending on whether the user is, for example, a professional, paraprofessional, or layman, francophone, anglophone, bilingual, or multilingual, and an urban, suburban, rural, or northern resident. For certain users, sophisticated technologies such as satellites or, in the future, glass fibre optics, may or may not have a noticeable effect on the habitability of a telecommunications network. The network may be habitable to the extent that these users also have hard-copy access in a conventional library--unless libraries are replaced by habitable modern alternatives.

Clearly, such considerations call for refinements of the evaluation criteria and measures proposed within various behavicural research paradigms. We need more effective techniques for determining the precise nature and amount of, for example, a given user's satisfaction with given aspects of the system environment. Scientific and technical communication is in the final analysis a cognitive process; the evaluation of habitability in cognitive terms may have to include new, nontrivial criteria from psychology, sociology, anthropology, linguistics, and related areas. The concept of habitability would serve as a heuristic device for these refinements. Other applications of this concept would, of course, include (1) behavioural assessment of existing communication systems to improve their habitability, and (2) behavioural assessment of proposed innovation alternatives to determine their habitability tradeoffs, module options, and the like.

Strategies

Provided we can agree that the environmental viewpoint and, in particular, habitability merit further exploration, we shall need workable research strategies. Heaps and Ingram (1972) have recommended an orientation toward specific national goals as the key strategy for Canadian information science. A goal-oriented strategy is not inconsistent with the integrative behavioural research envisaged here. In fact, the proposed conceptual integration and refinement could not be accomplished without massive empirical feedback from real-world situations including problems of national importance. Among major impediments to information transfer in Canada, the problem of "local habitation" (note the environmental flavour), which involves communication behaviour at "local" levels, is thought to be especially serious in terms of such responses as "personal frustrations" (Heaps and Ingram ibid.).

It is clear that cooperation based on enlightened self-interest will be needed to cope with the ramifications of the suggested behavioural research. Voluntary cooperation among relevant industries, government programmes, and academic institutions is vital to this endeavour.

HABITABILITY

We shall need a coordinated set of activities. Specific tasks might include data base activities, long-range planning, case studies, conceptual integration and criteria refinement, testing, and applications to meet particular needs.

FEASIBLE PRELIMINARIES

The first of the tasks just mentioned--creation and analysis of a data base--is, in principle, feasible in the near future. I am exploring specific options with Tom Atkinson of York University's Institute for Behavioural Research. We are assuming that a data base covering documents and projects with in-depth specialization in scientific and technical communication behaviour would have considerable utility for the information science community, regardless of its ultimate use in the proposed integrative research. A behavioural data base ranging over the expanded topical scope envisaged by us is not available anywhere. The need for such specialized data collections is underlined in the Science Council (1969) document known as the "Katz Report." Because of the specialized focus, our data base effort would be relatively modest in terms of its input volume. It would, however, entail computer-based processing and cooperative input exchanges with related information-gathering efforts.

The data base would emphasize behavioural research priorities in Canada, including the man-machine interface problem. It would contain (1) titles, abstracts, and indexes of the relevant literature, and (2) inventories of relevant research in progress. Taxonomic and state-of-the-art analyses of the data base contents would be attempted as a prerequisite to the suggested research.

REFERENCES

CUADRA, C.A., ed., 1966-, Annual Review of Information Science and Technology: Chicago, Encyclopaedia Britannica/William Benton; Washington, D.C., American Society for Information Science.

FRASER, T.M., 1969, "Relative Habitability of Dwellings--a Conceptual View," Ekistiks, v. 27, no. 158, p. 15-18.

HEAPS, D.M. and INGRAM, W.D., 1971, "Computer Recognition and Graphical Reproduction of Patterns in Scientific and Technical Style," Proceedings of the American Society for Information Science, v. 8, p. 257-261.

_____, 1972, "Information Transfer in Canada," in: Piternick, A.B., ed., American Society for Information Science/Western Canada Chapter: Proceedings, Fourth Annual Meeting: Vancouver, University of British Columbia, p. 171-181.

IRONSIDE, D.J., 1971, "Some Thoughts on Uses of Data Banks," Canadian Library Journal, v. 28, no. 1, p. 2-7.

KING, D.W., 1968, "Design and Evaluation of Information Systems," in: Cuadra, C.A., ed., Annual Review of Information Science and Technology: Chicago, Encyclopaedia Britannica/William Benton, v. 3, p. 61-103.

MAUERHOFF, G.R., 1971, "Some Aspects of Canada's Highly Qualified Manpower Resources with Special Reference to the CAN/SDI Project," American Society for Information Science/Western Canada Chapter: Proceedings of the Annual Meeting: Calgary, University of Calgary, p. 11-23.

MOOERS, C.N., 1960, "Mooers' Law: or, Why Some Retrieval Systems Are Used and Others Are Not," American Documentation, v. 11, no. 3, ii.

PAISLEY, W.J., 1968, "Information Needs and Uses," in: Cuadra, C.A., ed., Annual Review of Information Science and Technology: Chicago, Encyclopaedia Britannica/William Benton, v. 3, p. 1-30.

PLATT, R.B., 1971, "Environment," McGraw-Hill Encyclopedia of Science and Technology: New York, McGraw-Hill, v. 5, p. 12-15.

Science Council of Canada, 1969, Report No. 6: A Policy for Scientific and Technical Information Dissemination: Ottawa, Queen's Printer.

Telecommission, 1971, Instant World: A Report on Telecommunications in Canada: Ottawa, Information Canada.

TRUSWELL, J.S., 1972, "The Development of a Northern Environment Information System," in: Piternick, A.B., ed., American Society for Information Science/Western Canada Chapter: Proceedings, Fourth Annual Meeting: Vancouver, University of British Columbia, p. 7-18.

WATT, W.C., 1968, "Habitability," American Documentation, v. 19, no. 3, p. 338-351.

WHITTAKER, R.H., 1971, "Ecosystem," McGraw-Hill Encyclopedia of Science and Technology: New York, McGraw-Hill, v.4, p. 449-453.

WOLTERS, P.H. and BROWN, J.E., 1971, "CAN/SDI System: User Reaction to a Computerized Information Retrieval System for Canadian Scientists and Technologists," Canadian Library Journal, v. 28, no. 1, p. 20-23.

FABS (FORMULATED ABSTRACTING): AN EXPERIMENT IN REGULARIZED CONTENT DESCRIPTION (FABS: UNE ETUDE POUR LA FORMALISATION DU RESUME SIGNALETIQUE)

Brian Harris
Linguistics Documentation Centre
&
Thomas R. Hofmann
Dept. of Linguistics & Modern Languages
University of Ottawa
Ottawa, Ontario K1N 6N5

ABSTRACT

A bilingual experiment is being conducted at the Linguistics Documentation Centre, University of Ottawa, into the elaboration of well structured formulary routines for making the writing of abstracts easier, at the same time standardizing and generally augmenting the information given in them. (A l'Informathèque de Linguistique de l'Université d'Ottawa, une expérience bilingue tente présentement, à travers un formulaire rigoureusement structuré, d'établir des procédés routiniers qui faciliteraient la rédaction des résumés signalétiques, tout en favorisant la standardisation et le nombre d'informations.)

CAVEAT

This is a preliminary report of research that has begun only recently (Nov. 1972). It is being conducted on a very small scale. Even so, not all of the work can be described within the space limit of this paper.

INTRODUCTION & PRINCIPLES

Harris (1971) has shown how traditional bibliographic descriptions -- the information and layout usually employed in bibliographies or on author-title catalogue cards -- are so regular that most of their syntax can be described by a formal context-free grammar. Some advantages of this regularity are:

(i) Interchangeability of catalogue cards and easy reading of other people's catalogues and bibliographies.

(ii) Clearly implied directions to the cataloguer/bibliographer as to what information to extract from documents; which vastly helps the maintenance of standards in the matter.

(iii) Fields that are well delimited for information retrieval systems.

In sharp contrast to this epitome of regularity stand content descriptions that are written in the form of running text: the typical example is abstracts. (By this definition we exclude from consideration uniterm, multiterm and facetted classification languages.) From the linguistic point of view abstracting remains freely creative; but consequently it lacks the advantages just mentioned.

Advantages (i) and (iii) above accrue to the user. So does (ii) eventually, but it first affects the documentalist. Those implied directions help the latter by telling him:

(a) What he must not fail to account for, even if it might not all appear of interest to him personally.

(b) The order in which to set it down, relieving him of many decisions about the expressions and punctuation to use.

Conversely, misunderstandings ommissions and garbling are reduced.

If the constraints of a well tried format make work easier, the corollary is that the operation should be faster and cheaper. Even if the format is not rigidly adhered to in practice, a training in it should habituate an intellect to seeking out requisite information and should leave suitable terminology ready to spring to mind to express it. The help that a tyro abstracter receives at present is usually vague general advice such as, "Pick out what is new" or "What is the thesis of the author?"

Working with and within a regulated framework does not render the task entirely mechanical. Nothing of value can be had without allowing the bibliographer to make judgements, and the primary judgement to be made is that of selecting salient information. Of course much research is going on into automating the selection process; however, the potential connection with automatic abstracting lies beyond the scope of the present experiment.

There is a spectrum of abstracts that ranges from the very brief 'indicative' abstract -- of which the abstract at the head of this paper is an example -- through the 'informative abstract' which "presents the conceptual content" of the document, to the critical review which may sometimes rise to the status of an important work in its own right.

FABS

Amongst the criteria laid down for the formulary to be used in this experiment were the following:

(i) It must be as general as possible in the sense that it can be applied to as many documents and areas as possible.

(ii) It must avoid areas of judgement that are open to wide disagreement. Value judgements are to be excluded, although they may be put into...

(iii) Additional comments which are allowed as a supplement to the formulated information in case the specifics of the formulary neglect some important aspect. Indeed these additional comments should provide the basis for expansion of the formulary itself.

(iv) For the same reason, the analyst should be allowed to use a term of his own within the formulary when he feels that none of those provided fits the document.

These criteria and a number of other considerations are dealt with at greater length in Hofmann (1972).

METHODOLOGY

As a first step, an initial formulary was drawn up in English. This primitive version is reproduced as Fig. 1. When it is compared with guidelines issued by a well-established abstracting agency (Fig. 2), it is seen to cover much the same ground, and seems to prescribe a fullness somewhere between that of the informative and the indicative abstract. The difference lies foremost in the help given to the abstracter, because at the same time as he is directed to seek information he is provided with ready-made terminology and syntax. We will turn later to consideration of the product.

By application of the guidelines quoted above, the formulary has been considerably expanded, both in its terminology and in its constituent structure. A recent English version is Fig. 3. The expansions come about when one or other of the two analysts working on the documents (L. Légaré and M. Gelbert) exceeds the formulary. They then discuss with a coordinator whether their addition is really necessary or whether they have failed to make full use of what the formulary already provides. If their addition is accepted as necessary, it is incorporated in an updated version. After five months of this trial and error, the formulary seems to have a chance of reaching satisfactory stability: that we would define as no further changes in the constituent structure and not more than one addition to the vocabulary per 1,000 documents analyzed.

FABS

Légaré is working on French documents, and while so doing is compiling a French translation of the formulary which will be published later.

As a rough and ready way of early evaluation, we have had the analysts do abstracts of documents that had already been abstracted in reputable journals, namely "Language and Automation" and "Language and Language Behavior Abstracts". Of course our analysts did not see the other abstracts before doing theirs. We then compared the items of information given in the paired abstracts. An example of this sort of comparison is Fig. 4.

Approximately 50 documents have so far been analyzed.

EXPANSION OF THE FORMULARY

The increased flexibility, subtlety and precision provided by the larger vocabulary in Fig. 3 is obvious. However, it also contains some deadwood in the shape of expressions that were in the original formulary (Fig. 1) but have not proved their worth. Sooner or later, when we have enough data to do it safely, we shall have to prune.

At this point an explanation of the notation used in Figs. 1 and 3 is called for.

{....} groups a set of terms out of which the abstractor may (in some cases must) chose one term each time be works through the formulary.

(....) surrounds sets that are optional. The abstractor should always consider whether they are applicable, but often they are not. Conversely, sets that are not in parentheses are obligatory.

[....] surrounds an explanation which is intended to guide the abstractor but is not for use in his text.

CAPITALS distinguish the literals, i.e. the terms themselves that are to be used.

Underlined lower-case letters are used for variables.

Variables: x, y, z for noun phrases, s for sentences, l for language names or types; / separates alternative variables.

The expansion -- a significantly increased complexity -- in the constituent structure reflects the complexity of the documents themselves and was demanded if one wanted to retain more than the very briefest of indicative-type information.

The sophistications are of three kinds:

(i) The split of the original sections II and III into II, III and IV so as to introduce certain information that is now provided for by III and take some of the functional load off II and IV.

(ii) The addition of several optional constituents: see the parentheses notation above. This gives more flexibility.

(iii) Footnoting, to accommodate the additional comments prescribed by the criteria. Footnotes can also be used for bibliographical references.

(iv) Most important perhaps, the introduction of the iterative mechanism and its notation, (see below).

ITERATION

A single straight pass through the formulary would still only permit a brief indicative abstract. Indeed one way to force brevity is to insist that it be used that way (cf. Fig. 5, no. 3).

For analysis in greater depth, however, it soon became clear that a way was needed to put in more information, and this without having to make the distinction between essential and marginal information that footnotes imply. On the other hand, the general aim of the project required that the syntax be kept formally simple and, as linguists would say, 'transparent' (i.e. marked overtly). The solution was to allow any amount of 'backtracking'; at the extreme one can do a complete 'da capo' from section IV to section I. So that instead of being restricted to the order

(i) I → II → III → IV

one can go

(ii) I → II → III
↳ II → III → IV

or

(iii) I → II → III → IV
↳ I → II
↳ II
↳ II → III → IV

and so on.

Each backtrack is marked by the special conjunction '&', or by a string of &s in which there is an additional & for each section backwards that one moves. Translated into this notation, the above three examples are rendered:

(i) I II III IV

(ii) I II III &&II III IV

(iii) I II III IV &&&&I II &II &II III IV

The form 'and' is reserved for normal intra-sentential use within a section.

Though this may sound formidable, we feel the product can be read without strain: see the examples in Fig. 5, where some help is given to the eye by paragraphing.

INTERIM CONCLUSION

So far as we are aware, no other research is being done in this direction. Our immediate aim is to arrive at a satisfactory formulary; that is to say one which is comfortable and helpful for the abstractor, readable and useful for researchers, easily parsed by computer. We are feeling our way within these constraints. Reduction in the cost and increase in the speed of abstracting are very important problem areas, but cannot be tackled until an ulterior stage of research and would require more substantial funding.

REFERENCES

HARRIS, B., (1971) "A Justification and a Suggestion for a Linguistics Documentation Language." Cahiers linguistiques d'Ottawa, no. 1 (Sept. 1971), p. 7-25.

HOFMANN, T.R. (1972), "A Content Bibliography". Unpublished.

FABS

Fig. 1: FIRST VERSION OF FABS FORMULARY

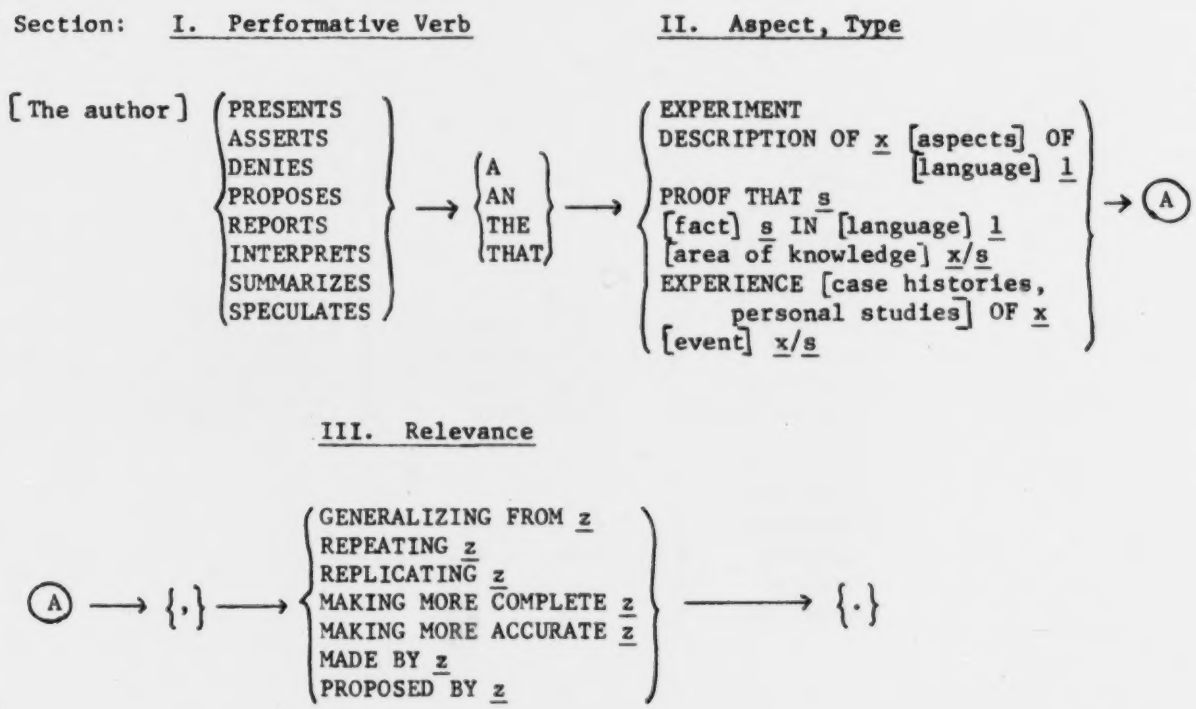


Fig. 2: TYPICAL GUIDELINES FOR ABSTRACTORS

Source: "LLBA"

INFORMATIVE ABSTRACT

An *informative abstract* presents the conceptual content of an article. Summarizing the essential ideas in an article, the abstract should answer the following questions:

1. What is the thesis of the author? What hypotheses or theories are presented?
2. How are the main hypotheses developed? What data are used? What methods of isolating data are used? Was novel methodology employed? Are the data qualitatively and/or quantitatively manipulated? What tests, scale, indexes, or other summarizing techniques are used?
3. What are the proofs or evidence relevant to the hypotheses?
4. What conclusions are drawn? Are the hypotheses, ideas, concepts, theories, etc., supported or rejected? What new relationships are found, old ones reaffirmed or rejected?

The *informative abstract* should show the meaningful, coherent relationship between the author's ideas and arguments; furthermore, it should enable the researcher to see the difference between one article and others on the same subject.

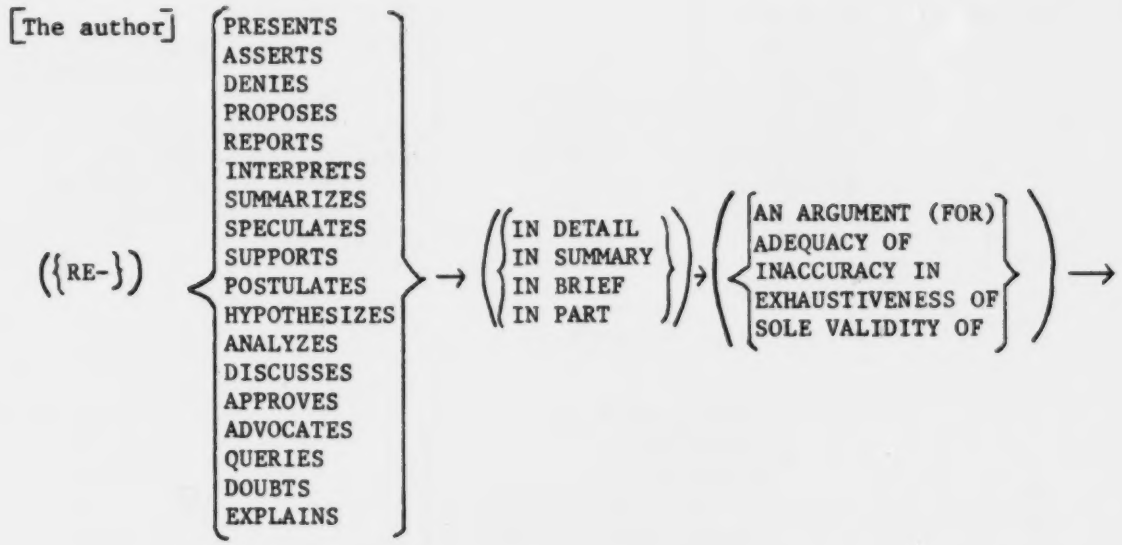
INDICATIVE ABSTRACT

Some articles (e.g., bibliographies, review articles, reports, and the like) cannot readily be summarized and require *indicative abstracts* which serve primarily as descriptive guides. An indicative abstract tells briefly what an article is about, what significant subjects it includes, and what its scope is.

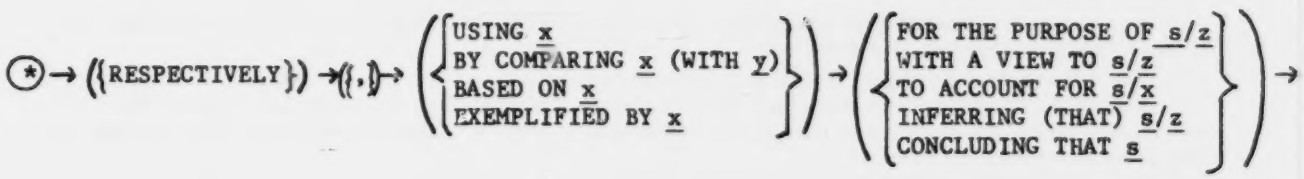
FABS

Fig. 3: FABS FORMULARY AT APRIL 1967

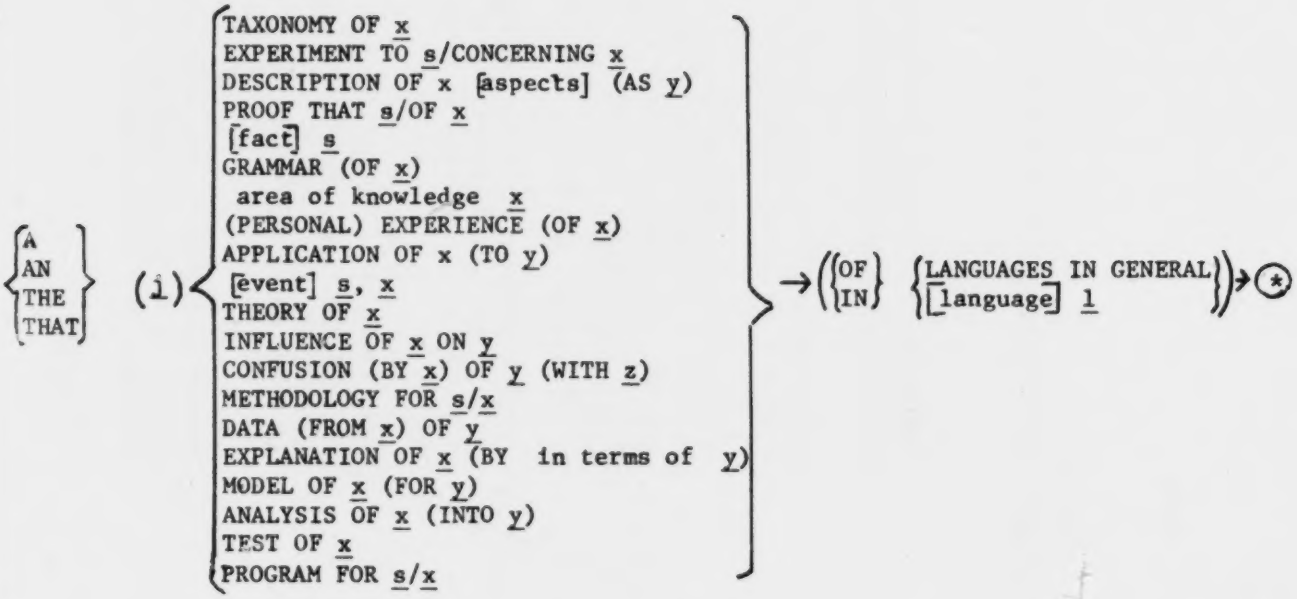
Section: I. Performative Verb



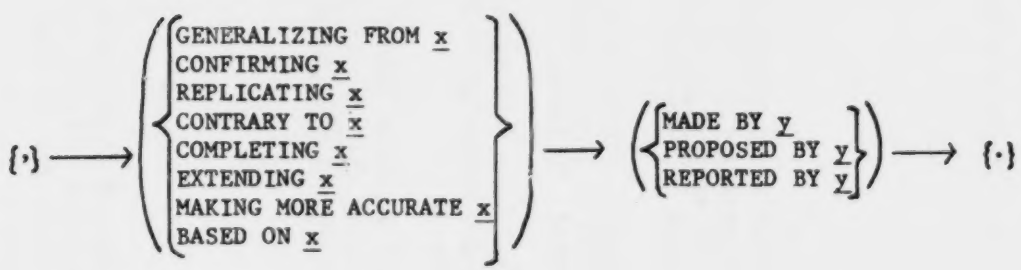
III. Means/Aims



II. Aspect/Type



IV. Relevance, Antecedents



FABS

Fig. 4

Martin, E., "Truth and Translation", Philosophical Studies, v. 23 (1972), p. 125-130

FABS abstract

ASSERTS DESCRIPTION OF translation AS pairing of every sentence of one language with a sentence of another so as to maximize conditions of truth and falsity

&&&& DOUBTS classical deductive truth-theory EXPLANATION OF human semantic competence, CONTRARY TO Tarski

&&&& ADVOCATES rule-governed THEORY OF semantic competence GENERALIZING FROM generative syntax

&&&& ASSERTS THAT Tarskian THEORY OF truth is essentially the translation from object language to metalanguage it must contain.

(67 words)

Note: Expressions underlined have been lifted directly from the original text. 41 words are the same in both abstracts.

"Language and Language Behavior Abstracts"

[Our paragraphing and underlining - BH/TRH]

A discussion of the value of translation in framing a theory of meaning.

Production of a theory of truth appears to be at least a minimal demand in explaining human semantic competence although an extension of Tarski's classical truth theory would provide such a theory, it is not necessary that a standard deductive theory be used.

A theory can explain human semantic competence if it appeals to rules which produce desired instances of (T) and can be mastered by human beings.

Translation works by pairing every sentence of one language with a sentence of another so as to maximize conditions of truth and falsity, and gives us an understanding of the semantic roles of the sentence part of each language.

Translations between two languages essentially provide us with translations between an object language and the metalanguage it must contain.

(139 words)

FABS

Fig. 5: EXAMPLES OF FABS ABSTRACTS

1. Paillet, J.P. & Hofmann, T.R., "Assumptions of Integrative Semantics", in Integrative Semantics I (SRG Monographs), 1972.

   PRESENTS IN BRIEF an ANALYSIS OF semantics INTO contents of messages and contents of lexical items and praxis phenomena

   &&&& PRESENTS a formal DEFINITION of a C-net AS an oriented graph of relationships between labelled nodes (semantic atoms) and unlabelled nodes standing for individuals to be eventually related to a referent in a universe, FOR THE PURPOSE OF representing the content structures of discourse

   &&&& POSTULATES the ANALYSIS OF lexical items INTO subnets of C-nets

   &&&& PROPOSES a DEFINITION OF understanding AS successful integration of the current segment of a discourse into the C-net already built up.

2. De Possel, R., "Les Résultats obtenus depuis fin 1968 en reconnaissance des formes et en particulier en lecture automatique par le R.A.M.I.", T.A. Informations, 1972, no. 2, p. 22-24.

   REPORTS IN BRIEF[1] RESULTS FROM an automatic reader OF roman characters USING a contextual (digraph and trigraph) and semi-sequential model FOR PURPOSE OF reading print and computer-output microfilm.

   [1] Full text to appear in Automatisme.

3. Moinfar, D., "Défini et non-défini en persan", T.A. Informations, 1972, no. 2, p. 20-21.

   PRESENTS MODEL OF recognition of definite and indefinite functional values IN Persian USING a table which relates marker and content.

4. Vaissière, J., "Contribution à la synthèse par règles du français", T.A. Informations, 1972, no. 2, p. 1-16.

   PRESENTS IN SUMMARY[1] a GRAMMAR OF & a PROGRAM FOR generation of prosodic features IN French BASED ON a phrase-structure syntax[2] & ON number of syllables per phrase, FOR PURPOSE OF speech synthesis.

   [1] of A.'s Thesis

   [2] that of CETA, Grenoble.

I.F. THESAURUS OF BUILDING SCIENCE AND TECHNOLOGY

(Thésaurus I.F. - Sciences et Technologie du Bâtiment)

Colin H. Davidson
Faculté de l'Aménagement
Université de Montréal

ABSTRACT

The I.F. Team have been struggling with the problems of information storage and retrieval in building, to the point that the basic tools had to be developed. These tools had to be explicit in their structure and construction, to compensate for the diversity of the building industry audience. The first of these tools is a hierarchical thesaurus, distinguished by its structure (a semi-lattice with nine levels) and its construction (based on a set of logical propositions). Candidate terms are treated in pairs, using a set of questions to ascertain systematically what are the relationships between them. Data processing helped with the construction of the thesaurus, particularly in terms of methodology. There is a scope for further work, particularly in preparing the bilingual (English-French) version of the thesaurus, now in draft.
(L'équipe I.F. a lutté avec les problèmes de l'enregistrement et de la recherche de l'information dans le domaine de la construction; fort de cette expérience, les membres de l'équipe se sont rendu compte qu'il fallait développer les outils de base. Ces outils devraient avoir une structure et une construction claires, afin de compenser la diversité de l'audience; les membres de l'industrie du bâtiment. Le premier de ces outils est un thésaurus hiérarchique, sa structure (un semi-réseau à 9 niveaux) et sa construction (basée sur une série de propositions logiques) représentent des innovations. Des termes candidats sont traités par paires, en se servant d'une série de questions afin d'identifier systématiquement la nature des liens entre les termes. L'informatique a aidé avec la construction du thésaurus, surtout en ce qui concerne la méthodologie. Il y a de grandes possibilités de continuer le travail, surtout pour la préparation d'un thésaurus bilingue (anglais-français) actuellement à l'étude).

ACKNOWLEDGEMENTS

The following people were involved with the project described in this paper: Michel Jullien (chief researcher), Pierre Garneau (responsible for data processing), Roger Camous, Claude Diacon, John Roberts, Leonard Wert.

I.F. THESAURUS

INTRODUCTION

This paper describes work that has been carried out by the "I.F. Team" at the Faculté de l'Aménagement, Université de Montréal. For reasons that are described below, we decided to develop a Thesaurus in the domain of building science and technology; for reasons that are also explained below, this lead to a number of special problems, the solutions to which - we believe - are of general significance for Thesaurus construction.

THE CONTEXT

Members of our team had been struggling with problems of information storage and retrieval in building for a while; the frustrations of trying to use the existing classification systems (devised to deal with bricks and mortar and quite unsuited for environmental phenomena or building processes) need not be described here. In addition, we were publishing a quarterly magazine (called Industrialization Forum (Ref.1) - whence our name "I.F. Team") which had a built-in coordinate indexing system - complete, but for a controlled vocabulary.

The nearest thesaurus was E.J.C. (Ref.2), but attempts to use it showed that (i) it did not cover our domain adequately, and (ii) it seemed that its structure was not appropriate - particularly in its liberal use of the related term (RT) cross references. A preliminary enquiry showed that one other English language thesaurus existed in the domain of building, - a loosely structured working document prepared by the Building Research Station in the U.K. (Ref.3); two were about to be prepared - one concerning building products in Canada (Ref.4) and a facetted thesaurus being considered in the U.K. (Ref.5). A French thesaurus, dealing with construction technology, had also been drafted (Ref.6).

THE PROBLEM

We were concerned with a domain of activity (building - science and technology) and not a "discipline"; an immediate consequence of this was that the people in the domain of activity have varied backgrounds, functions, vocabularies... to the point that communications were already difficult because of varied usages of terms. As a result, we were going to have to use every conceivable device - in the construction and presentation of our Thesaurus - to help standardize the use of terms.

THE I.F. THESAURUS

A thesaurus is a controlled and dynamic vocabulary, which covers a particular domain of knowledge and activities adequately. A thesaurus can be characterized by:

(i) its domain:

- particular domain,
- neighboring domains,
- degree of depth of coverage of these domains,
- extent of coverage of these domains;

(ii) its structure:

- internal structure,
- compatibility with other thesauri;

(iii) its different functional parts, (figs. 1 and 2);

(iv) its physical form;

(v) its method of use;

(vi) its process of evolution.

The I.F. Thesaurus can be characterized against each of these criteria. However, this paper emphasizes aspects of: (i) its structure, and: (ii) its process of evolution.

The Structure.

Because of the nature of the building industry "public" (varied, with no natural standardization of terms), we felt the need to adopt a "scientific" approach to structuring and constructing the thesaurus. By "scientific" we meant that there should be clear rules against which any and all decisions could be made; these same rules could be referred to in any subsequent arguement about the outcome of these decisions. The more rigorous the scientific approach, the better we would be compensating for the "undisciplined" (pluri-disciplined?) nature of our public.

The structure we adopted is rigorously hierarchical. Each descriptor has a defined level in the hierarchy (except for some general terms (Ref.7))and each descriptor has relationships to descriptors one level "up", one level "down" or on the same level in the hierarchy. The structure as a whole (a semi-lattice, with SCIENCES and ENVIRONMENT at the top level - level 1) is built up by systematic accretion of these individual relationships.

The relationships (which are registered by cross-references) were selected to avoid some of the uncertainties that prevail in many existing thesauri (Ref.8). These are: Broader Term (BT) and its reciprocal Narrower Term (NT) - restricted to the meaning: "type of"; Whole Term (WT) and its reciprocal Part Term (PT) - restricted to the meaning: "part of"; Related Term (RT) - restricted to the meaning: "associated with". The BT/NT and WT/PT relationships exist - and only

EXPERIMENTAL PSYCHOLOGY ***
...USE PSYCHOLOGY
EXPERIMENTS *0*
EXPERTISE ***
...USE SKILL
EXPLOSIONS *6*
EXPLOSIVES *6*
EXPONENTS *5*
EXPOSED CEILINGS *7*
EXPOSED CONCRETE FINISHE *7*
EXPOSURE *0*
EXPRESSION *0*
EXTERIOR LIGHTING *6*
EXTERIOR SPACES *5*
EXTERNAL ENVELOPES *6*
EXTERNAL FORCES ***
...USE LOADS(FORCES)
FIRE EXTINGUISHERS *8*
FIRE EXTINGUISHING EQUIP ***
...USE FIRE FIGHTING EQUIPMENT
EXTRACT VENTILATION *7*
EXTRACTION(CHEMISTRY) *6*
EXTRACTS(DOCUMENTATION) *6*
EXTRAPOLATION *5*
EXTRUDING *6*
EXTRUSIONS ***
...USE SECTIONS(MATERIALS)
EYE WINDOWS *7*
FABRICATION ***
...USE MANUFACTURING PROCESSES
FABRICATORS ***
...USE MANUFACTURERS
FABRICS *6*
FACADES *6*
LIGHTWEIGHT FACADES *7*
COORDINATING FACES *7*
FUNCTIONAL FACES *7*
FACETED CLASSIFICATION I *7*
TELECOMMUNICATIONS FACIL *5*
ADMINISTRATION FACILITIE *5*
PUBLIC SERVICE FACILITIE *5*
FACILITIES *4*
AGRICULTURAL FACILITIES *5*
CATERING FACILITIES *5*
CIVIL DEFENSE FACILITIES *5*
COMMERCIAL FACILITIES *5*
COMMUNITY FACILITIES *5*
CULTURAL FACILITIES *5*
DWELLING FACILITIES ***
...USE RESIDENTIAL FACILITIES
EDUCATIONAL FACILITIES *5*
ENTERTAINMENT FACILITIES ***
...USE RECREATIONAL FACILITIES
EXHIBITION FACILITIES *6*
HEALTH FACILITIES *5*
INDUSTRIAL FACILITIES *5*
MAINTENANCE FACILITIES ***
...USE STORAGE FACILITIES
MILITARY AIR FACILITIES ***
...USE MILITARY FACILITIES
MILITARY FACILITIES *5*
MOBILE FACILITIES *5*
MOTOR RACING FACILITIES *6*
MULTI-PURPOSE FACILITIES *5*
NAVAL FACILITIES ***
...USE MILITARY FACILITIES
OLYMPIC FACILITIES *6*
RECREATIONAL FACILITIES *5*
RELIGIOUS FACILITIES *5*
RESEARCH FACILITIES ***
...USE SCIENCE FACILITIES
RESIDENTIAL FACILITIES *5*
SCIENCE FACILITIES *5*
SPORTS FACILITIES *5*
STORAGE FACILITIES *5*
TEMPORARY FACILITIES *5*
TERMINAL FACILITIES *5*
TOILET FACILITIES ***
...USE TOILET SPACES
WELFARE FACILITIES ***
...USE HEALTH FACILITIES
BANKS(FACILITIES) *6*

FACING BRICKS *7*
FACINGS ***
...USE CLADDINGS
UTILIZATION FACTOR *6*
TONAL FACTOR METHOD *6*
FACTORIES ***
...USE INDUSTRIAL FACILITIES
FACTORS *0*
DANGER FACTORS ***
...USE HAZARDS
WEIGHTING FACTORS *0*
FADING ***
...USE DECREASE
FAIENCE *7*
FAILURES *0*
FAIR GROUNDS ***
...USE LUNA PARKS
FAIRS *6*
FALL-OUT SHELTERS ***
...USE ATOMIC SHELTERS
FALL(AUTUMN) ***
...USE AUTUMN
WATER FALLS *6*
FALSE DROPS *7*
FAMILIES *4*
FANLIGHTS *7*
FANS *8*
FAR EAST *7*
FARM HOUSES *6*
FARMS *6*
FASHION *4*
MODE(FASHION) ***
...USE FASHION
MECHANICAL FASTENERS *6*
BOLTS(FASTENERS) *7*
SPACERS(FASTENERS) *7*
FASTENING ***
...USE JOINING(PROCESS)
FAT CONCRETE *7*
FATIGUE LIFE ***
...USE FATIGUE RESISTANCE
FATIGUE LIMIT *7*
FATIGUE RESISTANCE *6*
FATIGUE STRENGTH AT N CY ***
...USE FATIGUE RESISTANCE
FATIGUE(MATERIALS) ***
...USE FATIGUE RESISTANCE
FAUCETS *8*
FEASIBILITY *0*
FEASIBILITY STUDIES *5*
GLACIAL FEATURES ***
...USE HYDROGRAPHIC FEATURES
HYDROGRAPHIC FEATURES *5*
TOPOGRAPHIC FEATURES *4*
FEDERAL SCOPE ***
...USE NATIONAL SCOPE
FEEDBACK *5*
NEGATIVE FEEDBACK *6*
POSITIVE FEEDBACK *6*
THRESHOLD OF FEELING(SOU *7*
FEELINGS ***
...SEE SCOPE NOTE
FELTS *6*
FENCES *7*
FENESTRATION *6*
FERMENTATION *6*
FERROUS METALS ***
...USE METALS
FERTILIZERS *7*
GLASS FIBER *6*
FIBER GLASS ***
...USE GLASS FIBER
FIBERS *6*
FIBROUS AGGREGATES *7*
FIELD TESTS *0*
FIELD THEORY(MATHS) *4*
ABELIAN FIELDS *5*
ELECTRIC FIELDS *6*
ELECTROMAGNETIC FIELDS *5*
ELECTROSTATIC FIELDS ***
...USE ELECTRIC FIELDS
MAGNETIC FIELDS *5*

Figure 1: Typical page of Alpha-Permuted Index.

* ENTHALPY *6*
BT thermodynamic properties
RT entropy
GRT energy

* ENTRANCES *7*
UF doorways
BT access spaces
RT fire escapes
porches

* ENTROPY *6*
BT thermodynamic properties
RT enthalpy
GRT energy

* ENVIRONMENT *1*
SN the complex aggregate of physical and geo-, bio- and socio-physical conditions that influence the life of organisms, individuals and communities.
...
use of a more specific term is recommended.
NT physical environment
socio-cultural environment
RT sciences
GRT interaction
relationships

* ENVIRONMENTAL ACOUSTICS *5*
UF architectural acoustics
BT acoustics
NT open air acoustics
room acoustics
PT acoustical defects
acoustical qualities
RT acoustic properties
acoustical control
acoustical environment
auditory perception

* ENVIRONMENTAL CONTROL *4*
UF space conditioning
NT acoustical control
air circulation control
air conditioning
hygrothermal control
luminous control
noise control
pollution control
WT building science
RT acoustical engineering
ambient environment
architectural design
environmental properties
illuminating engineering
thermal engineering

* ENVIRONMENTAL DESIGN *3*
NT architectural design
landscape design
urban design
RT building science
built environment
design

* ENVIRONMENTAL PERFORMANCE *5*
SN performance related to built environment and particularly to ambient environment.
NT acoustical performance
biological performance
hygrothermal performance
luminous performance
WT architectural design
environmental properties
industrial design
urban design
RT building performance
environmental requirements
GRT performance

* ENVIRONMENTAL PROPERTIES *4*
NT acoustic properties
biological properties
fire safety
hygrometric properties
optical properties
thermodynamic properties
PT environmental performance
RT environmental control
GRT properties

* ENVIRONMENTAL REQUIREMENTS *5*
SN in relation to ambient environment.
RT aesthetic requirements
architectural programming
design criteria
environmental performance
psychrometry
safety requirements
spatial requirements
GRT comfort
requirements

* ENVIRONMENTAL SCIENCE *3*
SN see also environment.
BT applied sciences
RT building science
built environment
ecology
psychology
sociology

* EPEIROGENY *5*
UF earth movements
orogeny
orography
WT geomorphology
seismology
RT earthquakes
hydrographic features
landforms
volcanology

* EPOXY *7*
BT thermosets

* equations(calculation) ***
USE FORMULAS(CALCULATION)

* EQUATORIAL REGIONS *6*
BT climatic regions
RT humid climates
tropical regions

* EQUILIBRIUM CONDITIONS *0*
GRT boundary conditions
conditions

* EQUILIBRIUM METHOD *6*
NT flexibility method
WT structural analysis
RT dynamic structural analysis
elastic analysis
plastic analysis
static structural analysis

* EQUIPMENT OPERATORS *5*
WT personnel

* EQUIPMENT(TOOLS) *4*
UF instruments
machinery
machines
tools
BT artifacts
resources
NT automated equipment(tools)
calculating equipment
construction equipment
draughting instruments
hand tools
machine tools
measuring instruments
observing instruments
power tools
production equipment
testing equipment
WT technology
RT ergonomics
manufactured products
processing
work(activity)

* EQUITY CAPITAL *6*
UF venture capital
BT capital(finance)

* EQUIVALENCE *0*

* erecting ***
USE ERECTION

* ERECTION *5*
UF erecting
tilting up
WT building technology
RT assembly processes
building processes
construction(techniques)

* ERGODIC THEORY *4*
WT mathematical analysis
RT probability theory

* ERGONOMICS *4*
UF man-machine interface
BT engineering
RT equipment(tools)

(E6) Figure 2: Typical page of Alpha-Hierarchical Index.

exist - between descriptors on one level in the hierarchy and appropriate descriptors one level "up" or one level "down" the hierarchy; The RT/RT relationships exist - and only exist - between descriptors on the same level of the hierarchy (Ref.9).

This set of relationships is shown graphically in fig.3 and typographically (as used in our alpha-hierarchical index) in fig.4. These "hierarchical groups" as has been stated - are joined by systematic accretion to form the whole structure or "conceptual space".

Before dwelling on the importance of this rigorous structure, we wish to describe the procedures that were used to generate the individual hierarchical groups. The key for this construction work is a set of so-called "logical propositions".

Obviously there are preliminary stages in which candidates are collected, their usage evaluated and their meanings agreed; once this has been done a preliminary sortation is made (which used certain computerized techniques)making clusters of candidates that apparently were likely to be related. At this moment, terms in these clusters were taken - two by two - and "processed" with the logical propositions (fig.5).

Let us consider, for example, some of the descriptors in the hierarchical group TOWNS (figs. 3 and 4).

(i) Let I be NEW TOWNS and J be TOWNS

P 1: NEW TOWNS can be a type of TOWNS - true
P 3: NEW TOWNS is always a type of TOWNS - true

therefore: NEW TOWNS is a NT of TOWNS

(ii) Let I be BUILT-UP AREAS and J be TOWNS

P 1: BUILT-UP AREAS can be a type of TOWNS - false
P 2: TOWNS can be a type of BUILT-UP AREAS - true
P 4: TOWNS is always a type of BUILT-UP AREAS - true

therefore: BUILT-UP AREAS is a BT of TOWNS

(iii) Let I be STREETS and J be TOWNS

P 1: STREETS can be a type of TOWNS - false
P 2: TOWNS can be a type of STREETS - false
P 5: STREETS can be an element, a sub-set, a sub-system or an aspect of TOWNS - true
P 7: if STREETS exist then TOWNS exist - true
P 8: if TOWNS exist then STREETS exist - true

therefore STREETS is a PT of TOWNS (PT_s meaning immediately part of it).

Figure 3: GRAPHIC DISPLAY OF THE HIERARCHY

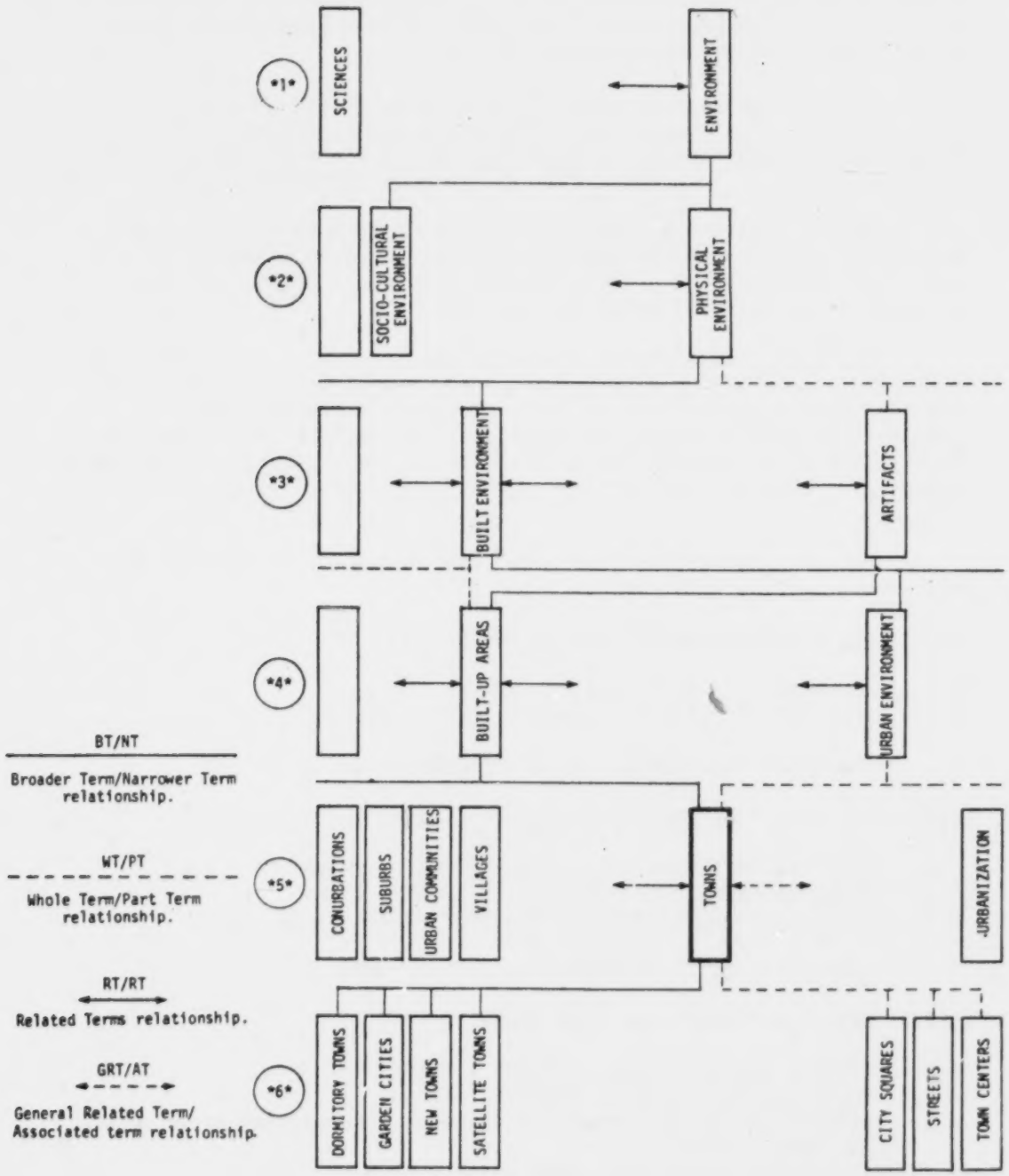
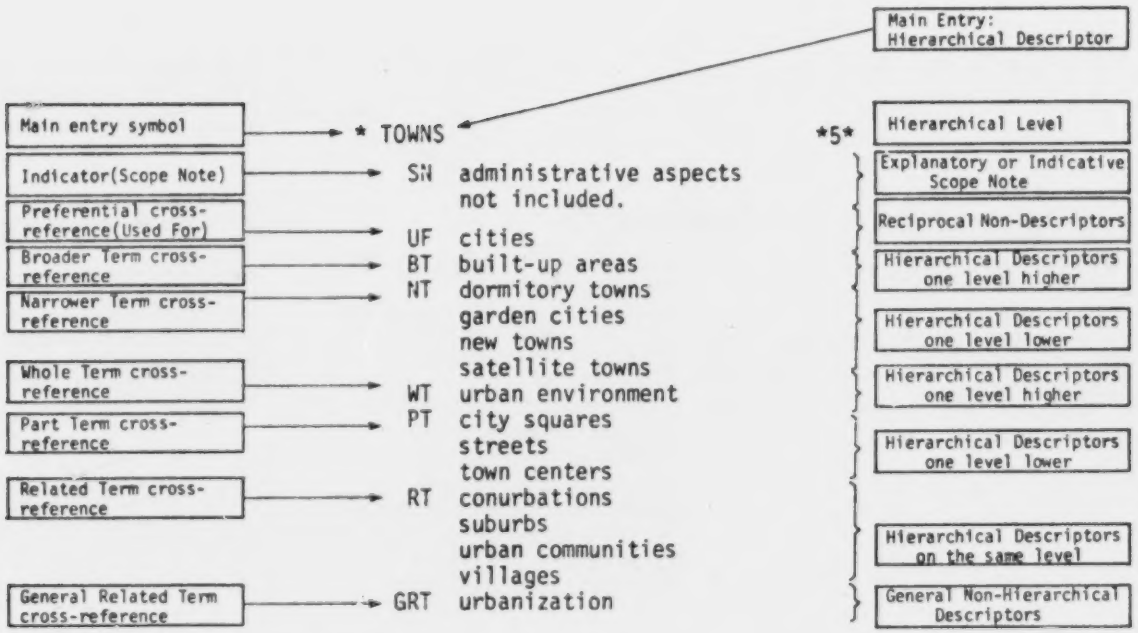


Figure 4: EXAMPLES OF MAIN ENTRIES IN THE ALPHA-HIERARCHICAL LIST

Hierarchical Descriptor

(This example is based on an existing group in the thesaurus, which has been completed, for the purposes of this section.)



Reciprocal Non-Descriptor

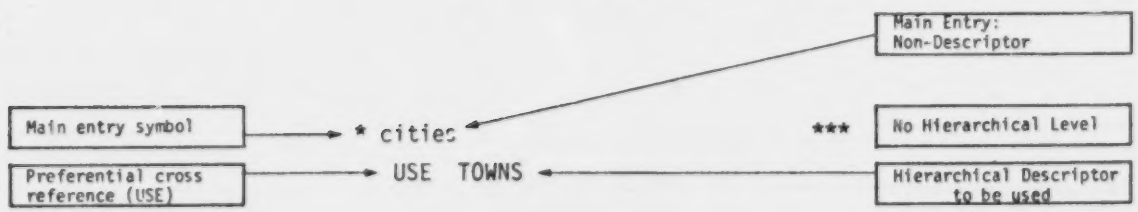
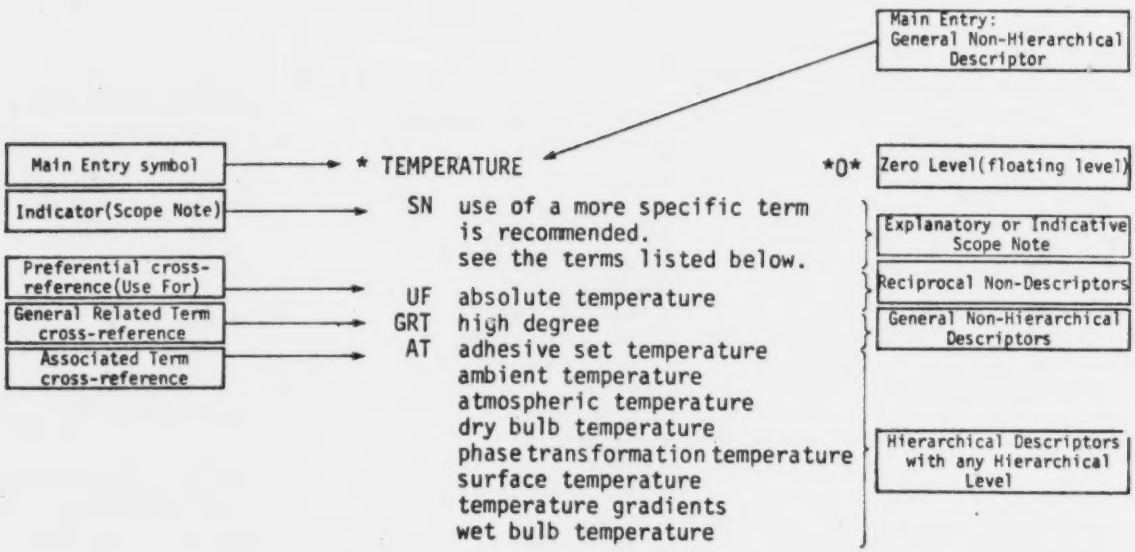


Figure 7: General Non-Hierarchical Descriptor



Guide Non-Descriptor

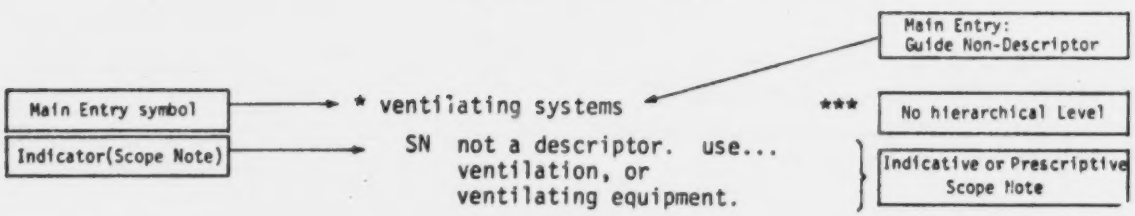


Figure 5: LOGICAL PROPOSITIONS

P1 (I) can be a type of (J).

P3 (I) is always of type of (J).

P2 (J) can be a type of (I).

P4 (J) is always of type of (I).

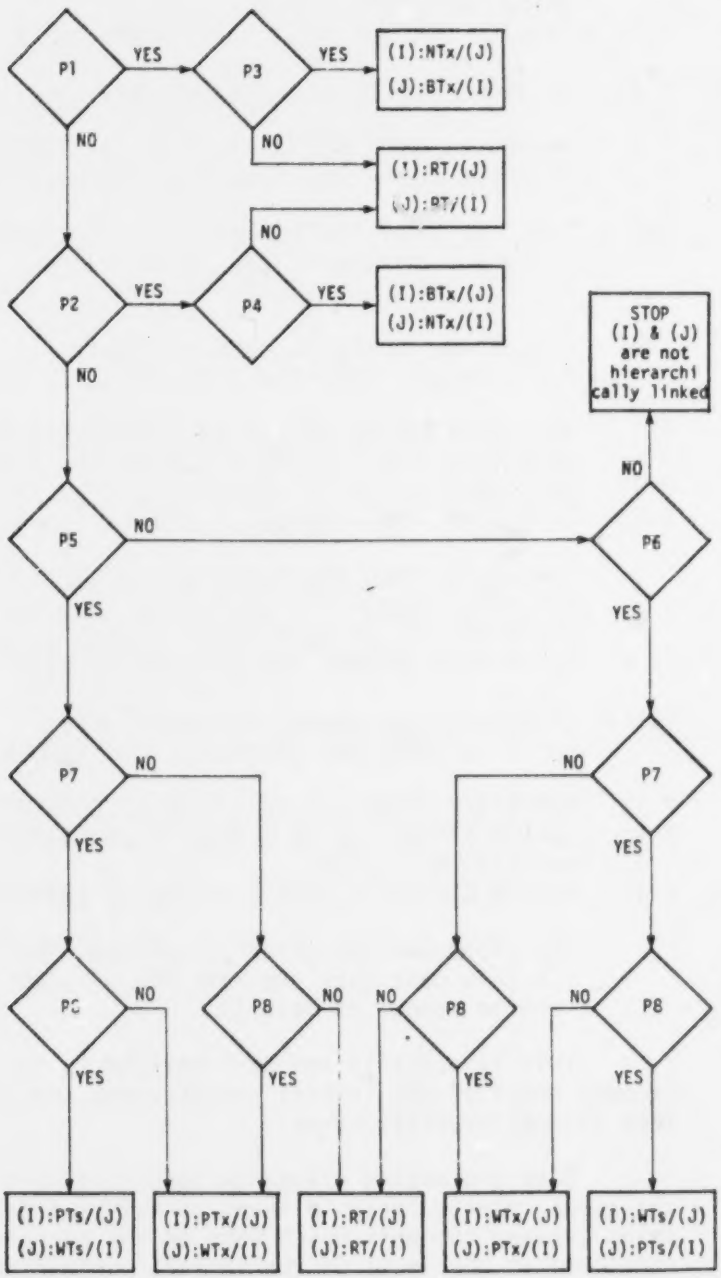
P5 (I) is an element, a sub-set a sub-system or an aspect of (J).

P6 (J) is an element, a sub-set, a sub-system or an aspect of (I).

P7 if (I) exist, then (J) exist.

P8 if (J) exist, then (I) exist.

(s : specific)
(x : less specific)



(iv) Let I be CITY SQUARES and J be TOWNS

P 1: CITY SQUARES can be a type of TOWNS - false
P 2: TOWNS can be a type of CITY SQUARES - false
P 5: CITY SQUARES can be an element, a sub-set, a sub-system or an aspect of TOWNS - true
P 7: if CITY SQUARES exist, then TOWNS exist - true (since TOWNS UF CITIES)
P 8: if TOWNS exist, then CITY SQUARES exist - false

therefore CITY SQUARES is a PT_x of TOWNS (though it is just possible there may be another intermediate term yet to be found)

(v) Let I be URBAN ENVIRONMENT and J be TOWNS
a similar procedure leads from P 1 to P 2 to P 5 to P 6 to P 7 to P 8; therefore URBAN ENVIRONMENT is a WT of TOWNS

(vi) Supposing, by chance, one had stated:
Let I be NEW TOWNS and J be SATELLITE TOWNS

P 1: NEW TOWNS can be a type of SATELLITE TOWNS - false
P 2: SATELLITE TOWNS can be a type of NEW TOWNS - false
P 5: NEW TOWNS can be an element (etc) of SATELLITE TOWNS - false
P 6: SATELLITE TOWNS can be an element (etc) of NEW TOWNS - false

therefore STOP! NEW TOWNS and SATELLITE TOWNS are not hierarchically related. In the event, they are both NT of TOWNS, and could be RT's of each other (though in practice it may not be worthwhile showing the RT relationship)

(vii) Supposing, by chance, one had stated:
Let I be DORMITORY TOWNS and J be GARDEN CITIES

P 1: DORMITORY TOWNS can be a type of GARDEN CITIES - false
P 2: GARDEN CITIES can be a type of DORMITORY TOWNS - after some hesitation: - true
P 4: GARDEN CITIES is always a type of DORMITORY TOWNS - false

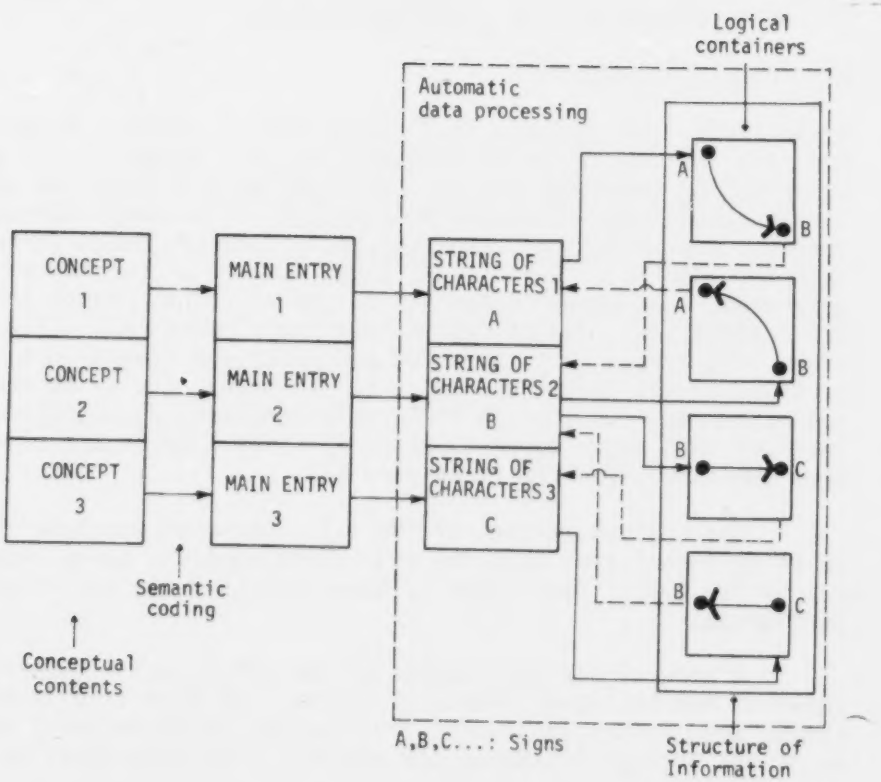
therefore GARDEN CITIES is a RT of DORMITORY TOWNS
(In this case they are both NTs of TOWNS, relationships which emerge sooner or later).

This is actually not as laborious as it seems, because one becomes adept at the logical propositions and only has to linger over them in controversial cases.

Data processing played an important part in the structure of the thesaurus, particularly in terms of methodology (fig. 6). The relationships between concepts which make up the hierarchy of the thesaurus are reflected at the level of the logical containers within the automatic data processing; the logical containers must be able to be processed

Figure 6: The role of data processing.

Suppose concept 1 is at level (n - 1), concept 2 is at level (n + 2), concept 3 is also at level (n + 2);
(i) Suppose that the logical propositions show that A is a BT of B - the data processing will (a) check that B is on one level below A and (b) then enter the cross reference and its reciprocal (B is an NT of A; A is a BT of B).
(ii) Suppose that the logical propositions show that C is an RT of B - the data processing will (a) check that C is on the same level as B and (b) then enter the cross reference and its reciprocal (B is an RT of C; C is an RT of B).



(2) See introduction 6. Thesaurus Development.

without knowing about the conceptual content to which they correspond.

Computer science does not have access to the semantic content of the data that is being processed; it deals only with the processing of symbols and numbers. Because of this characteristic, computer science provides - at all stages - the best possible logical check of the chosen structure and its rules. Identifying and dealing with special cases and "exceptions" have a detrimental influence on the clarity, concision and efficacy of the computer programs; any rule that has no logical relationship to the set of rules will be spotted quickly. In this way, computer processing of data is an effective way of appreciating the elegance and consistency of the structure as a whole.

What does this structure mean for the user of the Thesaurus? We have used the term "conceptual space" to describe the structure as a whole; each descriptor (which is a "label" for a concept) has an unique position within the conceptual space. Experience shows us that this is a great asset in guiding towards the best descriptor with which to describe a concept - whether in indexing comments, indexing queries or whether for other purposes for which concepts and their relationships are useful. The "conceptual space" is too complicated to represent graphically in its entirety (see, however, fig. 3 for an extract relating to our example TOWNS); the regular use of the Thesaurus - by constantly referring to the hierarchical groups - quickly forms a mental image of the main features of the "conceptual space".

The Process of Evolution.

A thesaurus - we stated - is a controlled and dynamic vocabulary... In our case, we developed the Thesaurus up to a certain point prior to using it for indexing; during that time, we drew upon the work of the Construction Industry Thesaurus team (Ref.5) whenever appropriate, to increase the chances of compatibility. Then the draft thesaurus was used for indexing. During this time, some candidate terms were being generated; they were "fitted in" to the structure, using the same logical propositions. Sometimes obvious "gaps" were thus filled in; sometimes term errors were corrected and sometimes (though not often) whole sections of the Thesaurus had to be "slipped" down or up one level (this had major repercussions on RT/RT relationships; in practice we found that any such major upheaval actually often lead to the elimination of some questionable RT/RT relationships).

The English edition of the I.F. Thesaurus was published as soon as we felt confident that the rate of changes was being reduced; this was in May 1972. Since then we have published one set of addenda and corrigenda.

A more significant aspect of the process of evolution is particular to the bilingual status of Canada. We have translated the Thesaurus into French (a "rough" translation, which reflects the structure of the English Thesaurus but contains some term forms we are not

satisfied with). This involved preparing sets of computer programs, enabling monolingual or bilingual alpha-hierarchical and non-hierarchical listings to be produced - with English or with French entry. Possibly thanks to the discipline in the structure, we did not encounter too much difficulty with the descriptors, though many of the non-descriptors did not need to be translated.

More significantly, however, is the fact that we are now collaborating with the Groupe Latin (a sub-group of Commission W.52 of the International Council for Building Research, Studies and Documentation) in preparing a basic latin-language thesaurus for building. Thanks to the collaboration of the researchers in the Groupe Latin, the coverage of the French Thesaurus is being improved; this in turn will be fed back into the English edition.

Obviously, the massive addition of new terms poses major problems - both in the hierarchical structuring of these terms and in data processing. An ideal we would like to attain would be to have the Thesaurus on-line on a CRT, so that one could "walk-through" the conceptual space, making improvements on the way.

CONCLUSIONS

This paper has scarcely mentioned the problems of indexing and the related problems of indexing rules. It is our experience, however, that for all aspects of working with, or on, a thesaurus the "scientific" approach is advisable. The rigorous structure enables the user to identify the position of the concept within the conceptual space; the logical propositions enable the hierarchical groups to be fully understood, and the significance of the relationships between terms to be grasped. Once this has been accomplished, indexing rules and search strategies can be built up appropriately.

Footnotes.

(1) Industrialization Forum - Building: Systems, Construction, Analysis, Research; published in English and in French jointly at Massachusetts Institute of Technology, Washington University and Université de Montréal; see: Wert, Leonard, "Information Retrieval and Industrialization Forum". Industrialization Forum, Vol.1, No.1 (October 1969), pp.11-17.

(2) Engineers Joint Council. Thesaurus of Engineering and Scientific Terms. New York, Engineers Joint Council, 1967, 690pp.

(3) Building Research Thesaurus; compiled in the Library of the Building Research Establishment, revised edition, Garston, Building Research Establishment, Department of the Environment, 1972, 1 vol.

(4) Canada. Department of Industry, Trade and Commerce. Materials Branch. Thesaurus of Canadian Construction Terminology, preliminary edition. Ottawa, Department of Industry, Trade and Commerce, 1971, 1 vol.

(5) Roberts, Michael, Chris. Eve, PeterLinn and Ellen MacHale: Construction Industry Thesaurus; second preliminary draft. North Western Polytechnic School of Librarianship and the Polytechnic of the South Bank, 1971, 1 vol.

(6) Thésaurus du Bâtiment et des Travaux Publics. Paris, Institut Technique du Bâtiment et des Travaux Publics, 1970, 149pp. This document is now being merged into a basic latin-language building thesaurus in preparation by the Groupe Latin of the International Council for Building Research, Studies and Documentation.

(7) General Terms exist ex-hierarchy (e.g. TEMPERATURE, see fig.7) they have associative relationships with other General Terms (designated General Related Term - GRT) or with hierarchical terms (designated Associated Term - AT). Since the General Terms are ex-hierarchy, there can be no question of hierarchical levels in their cross-references.

(8) For example, E.J.C. puts part terms in the RT groups; whereas the Thesaurus of Canadian Construction Terminology puts them in the NT relationships. For an attempt to sort out RT's, see the introduction to: Barhydt, Gordon C. Charles T. Schmidt and Kee T. Chang. Information Retrieval Thesaurus of Education Terms. Cleveland, Press of Case Western Reserve University, 1968, 133pp.

(9) This restriction means that the associative relationship is only shown between descriptors that are "comparable" - in the sense that they represent concepts of analogous generality or specificity.

EDUCATION FOR INFORMATION SCIENCE: AN APOLOGIA AND AN INTRODUCTION (L'EDUCATION POUR LES SCIENCES DE L'INFORMATION APOLOGIA ET UNE INTRODUCTION)

Professor C. David Batty
Graduate School of Library Science
McGill University, Montreal

ABSTRACT

Education for information science must at the same time cater for present needs and plan for future developments. It is necessary to examine the constituent elements of the discipline, and to build up a clear view of its scope and direction. There are several aspects to consider: the context of demand; the content; methods of education; applications. There are several categories of problems to consider also: first job knowledge versus long range flexibility; what is the core and what is the fringe; how far to mix theoretical and practical instruction and experience. (L'éducation pour les sciences de l'information doit répondre à la même fois aux besoins d'aujourdhui et aux espoirs de l'avenir. Il nous faut examiner les éléments qui constituent les sciences, et établir une répresentation exacte de leur champ d'action et de leur direction. Il nous faut considérer plusieurs aspects; le contexte du besoin; le contenu; les modes de l'instruction; les applications. Il y'a aussi plusieurs types de problèmes à considérer; les connaissances nécessaires au premier emploi versus les connaisances nécessaires pour la vie professionnelle; quels éléments constituent le centre des sciences de l'information et quels sont à la périphérie; comment équilibrer l'instruction théorique et l'expérience practique.)

Education in any field is liable to present a paradoxical image. It offers an outline and exemplar of the current state of the art, reflecting accepted content and direction, and the strengths and weaknesses of past knowledge and practice, but at the same time, associating itself with research and the development of new ideas in the field, it tries to become a model for the future so that the professionals it educates may function properly in a world to come. It may thus become at worst a plaything of professional prejudice, at best a focus for intelligent planning. Information science is new enough still to be torn by discussion and even faction over the scope and definition of the field itself, and not only the educationalists have argued over the constituent elements of what they profess.

The papers presented in this session are representative of several aspects of the education problem: the context of demand; the content of the discipline; the method and direction of education; and

the applications of what we learn. I should like to offer by way of introduction, some general thoughts about information science and its constituent elements as they affect education.

Much of the argument about exactly what constitutes information science is philosophical; we should do better as professionals to adopt the inductive reasoning that we use in so many of our techniques, and build up a definition (if indeed we need or want one) from observation of what practices and principles are sine qua non in our work. I offer no such analysis or synthesis but I feel it necessary to refer to some of the factors and arguments for what they may tell us.

Various schools of thought in the debate to define information science swing now to sociology, now to communication theory, now to computer science, now to administration - but all have a common (and hilariously suspect) factor in defining it explicitly away from library science. It is true that the librarian is traditionally concerned with libraries as storehouses of knowledge, offering documentary services to patrons: the loan of books; reference services based on books (often by directing the enquirer to the appropriate book) and collection building for the future. In this orientation to the physical aspects of information, although the problem-solving situation is still a common element, the librarian has attempted to maintain a passive role, and to promote only a helpful context in which the user may exploit the store of knowledge. But the growth of the quantity and complexity of knowledge makes it impossible now for any user in any community to be able adequately to comprehend or exploit even his own field of knowledge, even less the relevant ideas or methodologies in fields beyond his own. A dynamic role has now been defined, and is what concerns the information scientist.

But the attempts to draw these lines of demarcation may prove to have been a waste of time and energy; a study of the essential practices and the professional knowledge of information scientist and librarian reveals differences largely in the level and complexity of information handled, in the formality or informality of the mode of publication or presentation of the material handled, and in the degree of dynamic interaction with the patron. Structurally the nature of the transaction is fundamentally similar; a knowledge of information sources, of information coding and handling techniques, and of course, of the management of the information system, from its basic routines to its physical context and raison d'être. Certainly there are differences in orientation - but they constitute differences of degree rather than of kind, and of a degree rarely so great as to amount to a difference of kind.

A more fundamental difference distinguishes the scientific information worker who must have a knowledge of the substance as well

as the structure. How, and to what level, that knowledge must be achieved is difficult to say - but it should encompass the fundamental principles of the discipline, its concepts, terminology, and techniques. It may not be necessary for this information worker to have been in scientific research, but he must know enough to be a fully understanding, able and equal colleague. But at least no less should he have an authoritative expertise in information science.

In the face of a rapidly changing society and its technology we cannot expect to predict in concrete terms the context, tools and operations of information workers even twenty years from now - but some at least of our students now in universities can expect a working life of forty years after graduation. It follows that we should concentrate on problem-solving, decision-making abilities, together with the objectives of information work, so that the individual will be able to respond to any context or problem. At the same time we must realise that these same graduates are needed now and in the next few years in positions that exist now and for which they must receive some appropriate training as a foundation for experience and promotion.

The first dimension concerns objectives. What balance of first job skills and long range education should a program contain? Neither the student nor the initial employer will be happy with a neglect of practical expertise that will enable the professional to succeed and achieve responsibility and power; but the profession as a whole will suffer if he is crippled when he gets there by a neglect of fundamental principle. The solution may be in the discussion of the remaining dimensions. The second dimension concerns content: is there a core, what does it compose, and how much of it, and how much of fringe, supportive subjects should be included. There is a growing acceptance of an identifiable core: the resources, transfer mechanisms, management and context of information. But there is an increasing feeling that we cannot exclude disciplines that form part of these techniques: computer science, operations research, communications and the media. There is also the apprehension that information science has no substance in the conventional sense, and that somehow some elements must be sought to introduce it. But we may as easily say that information science has no substance in the conventional sense; it is a discipline that deals with structure - indeed, structure is its substance.

The third dimension concerns the educational process; how should we balance theoretical and practical studies? Ours is a complex discipline, and needs a theoretical framework within which all aspects may be comprehended easily. At the same time it is a profession of practitioners. Somehow the educational process must acknowledge this duality; can we imagine many other professions turning their new practitioners loose on the world without a period of rigorous practical training under strict and authoritative supervision?

We cannot expect to find answers to all these questions immediately, but hopefully we may begin to define the questions more clearly and sensibly, and a well defined question, as any information worker knows, is a long way towards the answer.

NATIONAL GOALS FOR THE EDUCATION AND TRAINING OF PERSONNEL FOR A NATIONAL SYSTEM FOR SCIENTIFIC AND TECHNOLOGICAL INFORMATION (OBJECTIFS NATIONAUX DE LA FORMATION DE PERSONNEL EN VUE D'UN SYSTEME NATIONAL D'INFORMATION SCIENTIFIQUE ET TECHNIQUE)

Dorothy H. Evans
ABSTI Secretariat
National Research Council of Canada
Ottawa, Ontario K1A 0R6

ABSTRACT

Effective utilization of scientific and technological resources for the achievement of national objectives is highly dependent upon a nation's ability to develop an effective system for the acquisition, storage, retrieval and dissemination of scientific, technological and socio-economic information. The successful planning and implementation of a Canadian national system for scientific and technological information will be highly dependent upon the available human resources. Manpower requirements will include those for individuals possessing a wide variety of educational backgrounds and experience. A concerted effort on the part of all Canadian schools of library and information science to provide interdisciplinary programs for the education and training of STI specialists will be required. Educators in these schools should also be prepared to accept the challenge of contributing significantly to the design and development of programs of instruction for STI users, and to the provision of programs of continuing education for practising STI specialists, and to the development of reliable methods for assessing the effectiveness of such programs. (L'exploitation efficace des ressources scientifiques et technologiques en vue de la réalisation d'objectifs nationaux dépend beaucoup de l'aptitude d'une nation à élaborer un système efficace d'acquisition, de stockage, de recherche et de diffusion de l'information scientifique, technologique et socio-économique. La planification et la mise en oeuvre réussies d'un système national canadien de documentation scientifique et technologique dépendra largement des ressources humaines disponibles. Il faudra un personnel compétent, y compris des individus possédant une formation et de l'expérience dans beaucoup de domaines différents. Il faudra également un effort concerté de la part de toutes les écoles canadiennes de bibliothéconomie et des sciences de l'information en vue d'offrir des programmes

interdisciplinaires pour la formation des spécialistes de l'IST. Les éducateurs de ces institutions devraient aussi être prêts à jouer un rôle important dans la conception et la mise au point des programmes d'instruction destinés aux usagers de l'IST et dans la préparation des programmes de formation professionnelle continue destinés aux spécialistes actifs de l'IST, ainsi que dans l'élaboration de méthodes sûres servant à évaluer l'efficacité de tels programmes.)

The awareness on the part of both highly developed and developing countries of the importance of scientific and technological research and development for sustained economic growth has led these nations to adopt policies designed to promote the effective utilization of their scientific and technological resources. The beneficial effects of science and technology in improving the quality of life have been well demonstrated, as well as the serious undesirable consequences which ensue when industrialized countries fail to exercise prudence in the use of technology. National policy makers have come to recognize their increasing dependence upon the ability of experts to forecast not only the technological consequences but also the social consequences of the successful exploitation of technological innovations. The formulation of national policy in many areas is thus highly dependent upon the availability of relevant and timely scientific, technological and socio-economic information obtained from both national and international sources.

Users of scientific and technological information include not only scientists and engineers but also policy makers, planners and managers both in government and industry, as well as the general public. The primary goals of national STI systems are the effective utilization of accumulated knowledge in science, technology and related fields in order to achieve national objectives for the benefit of society; the promotion of the development of science and technology and the production of new knowledge; the provision of adequate information for management and policy makers in both the public and private sectors; and the focussing of the attention of all interested individuals and organizations on problems relating to the availability and use of information. Although much of the information required may be national in origin, it is obviously in the self-interest of every nation to gain access to the world's output of basic scientific knowledge and that portion of technological information which is applicable to the national R & D effort.

The acquisition, organization, storage, retrieval, and dissemination of the total world output of scientific and technological information is an undertaking too vast in scope for any one institution, agency or even any one nation to undertake. Provision for access to a

substantial portion of this information on a national scale requires the cooperation of many institutions, agencies, libraries and information centres scattered throughout the country. Arrangements for providing access to additional information can be made possible through various types of international exchanges both formal and informal.

The successful planning and implementation of a comprehensive national system for scientific and technological information in Canada will be highly dependent upon the human resources which can be effectively directed towards contributing in a significant way to the realization of that goal. Manpower requirements will include those for individuals skilled in library and information systems analysis and design; individuals possessing a thorough understanding of the design, operation and effective utilization of various types of systems for the storage and retrieval of information in and about the fields of science and technology; information specialists possessing a high level of familiarity with the use of conventional primary, secondary and tertiary sources as important access tools to specific types of information; subject specialists possessing a substantial background of knowledge in specific subject areas of science and technology; information technologists skilled in the use of specialized equipment for information processing, for the storage of documents in microform, and for their reproduction using a wide variety of reprographic techniques; as well as specialists in management sciences who will hopefully have had some opportunity to acquire, in a formal educational setting, an understanding of the problems associated with the provision of information services in libraries and information centres, and through the mechanism of a highly developed national system for the dissemination of STI, having as its focal point a national science library or national centre for scientific and technological information. Many of the individuals who will play important roles will have received most, if not all, of their formal education in institutions other than schools of library and information science, while others, although not having completed a formal program of professional education in library or information science, will hopefully have participated to some extent in the education and training programs offered by these professional schools.

Effective user-system interaction requires that the information specialist, whose education and training will be based primarily in the professional schools of library and information science, and whose role will be that of serving as an intermediary between the system and the user, possess a thorough understanding of the organization of information sources within the system, as well as the ability to match these information sources to the true information needs of the user. Effective performance in such a role requires that the information specialist be well informed concerning the various subject areas of the fields of science and technology and their interrelationships; in addition, he must be able to interpret the information needs of the individual user, in the context of the scientific or technological community or other organization of which he is a member.

EDUCATION OF STI PERSONNEL

Optimal utilization of the total information system available to the potential user is largely dependent upon his knowledge of the system, its capability, the effort he must expend in order to gain access to it, and the cost-effectiveness of the system judged in relation to his actual information needs. For many individual users or groups of users, the most effective method of introduction to the system would be through participation in well-designed programs of user education presented by skilled and experienced instructors and scheduled, in respect of time and location, with special regard to the convenience of users. Educators and students in the area of library and information science should give serious consideration to the possibility of their making significant contributions to the design and implementation of such user education and training programs, and to the development of appropriate techniques for the evaluation of their effectiveness.

The essential components of education and training programs for STI specialists, whose basic professional training will be obtained in schools of library and information science are: (1) a study of scientific and technological communities and their information needs; (2) a study of the organization and growth of knowledge and the relationship of science and technology to other areas of knowledge; (3) a study of the sources of scientific and technological information and (4) a study of techniques by which the information needs of scientific and technological communities may be met. A detailed outline of the topics which might be included in such a program has been embodied in a special report prepared by the Secretariat of the Advisory Board on Scientific and Technological Information, and distributed to the schools of library and information science in Canada. Studies relevant to many of the topics suggested might however be more profitably based in departments and schools of universities other than the schools of library and information science and might be made available to interested students in other departments of the university.

Since the efficient operation of an organism as complex in design and function as a national system for scientific and technological information requires the utilization of a wide variety of human resources, it is important that educational programs for personnel who may be involved in the operation contain a significant interdisciplinary component. Precautions should be taken to ensure that the various types of specialists involved have sufficient knowledge to permit them not only to perform their own role in an adequate fashion, but also to possess an understanding of the roles of other personnel, so that they may have a proper appreciation for the contribution of each individual to the proper functioning of the system.

Although many professional schools of library and information science in North America have in the past, provided and still continue to provide excellent education and training programs, containing elements which are not only relevant but indeed essential to the professional

preparation of individuals planning careers in the field of scientific and technological information, only rarely has any school attempted to provide a comprehensive program of studies, especially designed for the production of STI specialists, that is, those individuals who through their education, training and experience, possess not only a thorough familiarity with the literature of the many and varied fields of science and technology, and the many secondary services which provide bibliographic access to the primary literature, but also a comprehensive knowledge of the principal research and development agencies and institutions, which not only constitute the largest group of users, but also the largest group of producers of scientific and technological information.

However, the importance of STI was widely recognized, as is evidenced by the extensive activity which developed in several areas of research relevant to STI. The primary objective of much of this research was the development of more efficient and effective systems for establishing adequate bibliographic control over, and a wider and, at the same time, a more selective dissemination of, the voluminous output of scientific and technical literature which accompanied the worldwide expansion of research and development activities in the field of science and technology. Although this research, aimed primarily at improving methods for the storage and retrieval of information, was accompanied by numerous, heavily funded and unusually productive and illuminating studies of the role which the diffusion of scientific and technological information plays in contributing to the expansion of scientific knowledge, and the process of technology transfer, probably the main consequence of these research programs was the production of experts in extremely specialized areas of scientific and technological information and scientific and technological communication.

Generally speaking, there has been an absence of broad scale application of the findings of the majority of these investigations to the design of economically viable information systems and services to meet the needs of a wide variety of scientific and technological communities, in a manner designed to permit the optimal utilization of information resources, with the aim of increasing scientific productivity, technical innovation, economic growth, and thus the attainment of more general national and international goals. Thus, there is still much reason to believe that we have not yet reached the necessary level of knowledge and understanding of the true value of information, that is, the right information, relevant to human needs; nor have we been able to formulate policies for ensuring that such information is made available, in a useful form, to all who need it.

The ever-increasing volume of scientific and technological literature, as a result of the emergence of new areas of scientific and technological knowledge, many of a highly interdisciplinary nature, together with the shifting research interests and new information requirements of many individuals whose professional and organizational responsibilities

are becoming more frequently subjected to rapid change, has produced a continuously increasing demand on information-providing facilities, be they industrial libraries or information centres, university libraries, or information service units for individuals employed by governmental agencies at local, provincial and federal levels. The achievement of self-sufficiency of any type of library or information centre in meeting the information needs of its primary clientele is thus becoming virtually impossible. Consequently, there has resulted an increasing need for interlibrary cooperation. In order to make such cooperation economically and functionally feasible, it is necessary to aim at a higher level of coordination of all activities related to library operations and the provision of information services, than that prevailing in the past.

The application of computer-based techniques to many areas of information handling has resulted in an increased need for cooperation in the establishment of standard methods for the creation and use of machine-readable bibliographic records. Resource-sharing among libraries is largely dependent upon the availability of continuously updated access tools, such as national union catalogues, union lists of serials, and directories of information sources, employing standardized formats. In addition, there is an increased requirement to utilize in a more effective manner than in the past, human resources such as scientific and technical personnel possessing a high level of expertise in highly specific subject areas.

Recent years have witnessed an increasing degree of international cooperation in all aspects of the provision of library and information services, with the goal of achieving greater levels of compatibility between the information systems currently being operated under national control or sponsorship, or by national professional societies and commercial organizations.

International organizations such as Unesco, the International Council of Scientific Unions, the International Federation for Documentation, the Organization for Economic Cooperation and Development, and the International Federation of Library Associations, are becoming increasingly effective agencies for promoting the coordination of the development of national information systems and the international exchange of scientific and technological information.

Educators in schools of library and information science have a very serious responsibility to bring to the attention of their students, and to awaken in them an interest in, the many attempts at cooperation and resource-sharing which are beginning to play an extremely important role in the formulation of national information policies and the promotion of international cooperation.

Moreover, professional schools of library and information science should endeavour to provide for their students a realistic assessment of

their future responsibilities, and to create within them a true sense of professionalism by providing for their students and thus for all future practitioners, an education based upon a firm foundation of knowledge of the science, art and practice of their profession.

It is unlikely that the challenge of meeting the future national and international needs for effective information services will be met except through the continuing efforts of individuals who are willing to commit themselves to a lifetime program of learning. Schools of library and information science can play an important role by providing in cooperation with professional associations such as the Canadian Association for Information Science and the Canadian Library Association, opportunities for continued personal and professional growth through the planning and implementation of readily accessible and well-designed programs of continuing education.

In view of the high level of competence and degree of specialization that will be required of information specialists who will enter the field of scientific and technological information, it seems highly unlikely that the future national manpower requirements will be met, unless the schools of library and information science are willing to incorporate in their curricula, provision for specialized educational programs for future STI specialists. At the same time, the necessity for providing their students with a firm foundation in the fundamentals of librarianship should not be overlooked. The development of adequate programs will require an extensive cooperative effort on the part of all Canadian schools of library and information science, so that economically feasible programs for STI specialists may be made available to interested, qualified students throughout all of Canada. The successful implementation of an effective and efficient national system for scientific and technological information will require well trained, highly motivated and highly qualified information specialists to man the many libraries and information centres scattered throughout Canada, without whose cooperation no truly national system can ever become operational.

DATA - THE MISSING LINK IN THE EVOLUTION OF LIBRARY SCIENCE CURRICULA (LES DONNES - L'ANNEAU MANQUANT DANS L'EVOLUTION DES PROGRAMMES D'ETUDES EN BIBLIOTHECONOMIE)

Frank T. Dolan
School of Library and Information Science
The University of Western Ontario, London, Ontario N6A 3K7

ABSTRACT

The role of the library is to serve as a catalyst in enabling the user to interact directly and optimally with the body of recorded knowledge. The role of the Library School is to prepare individuals to operate as information professionals inside or outside the institutional framework. Data provides the foundation on which all research and most publication is built. Yet documents and their handling still dominate most Library School curricula and data management principles, practices and systems are not covered satisfactorily. There are five important reasons why this situation must change if libraries and Library Schools are to fulfill their mandates: 1) The existence of machine-manipulable data bases in practically every discipline. 2) The existence of "hard data" bases in many areas. 3) The trend toward computer-aided, indeed computer-based publication and dissemination. 4) The trend toward data rather than document transfer. 5) The important role of administrative data-processing within a library to control the operating systems. The technology required for data management is state-of-the-art. Generalized Data Management Systems provide the support for real breakthroughs in data handling in libraries and information centers. As an example of the simplicity, the flexibility and the power of these systems, one such system, DATAMAN, is applied to a circulation data base. An understanding of data management principles, practices and systems is essential for the modern librarian. Courses covering this topic must be made an integral part of Library School Curricula. (Le rôle d'une bibliothèque est de servir de catalyseur en aidant l'usager à garder le meilleur contact direct qui soir possible avec la somme de connaisances qui nous est transmise. La fonction primordial de l'école de bibliothéconomie est de préparer les candidats à devenir

spécialistes en information, à l'intérieur ou à l'extérieur des cadres institutionnels établis. Les données constituent le fondement essentiel de toute recherche ou de publication. Les programmes d'études de la plupart des écoles de bibliothéconomie restent toujours fortement axés sur le traitement et la circulation des documents. Les principes de gestion des données, leur application pratique, et l'ensemble des systèmes qu'ils engendrent ne reçoivent pas une attention satisfaisante. Certaines raisons importantes peuvent être invoquées pour justifier le changement d'attitude requis afin que les bibliothèques remplissent leur rôle. 1) L'existence des bases de données bibliographique qui peuvent être traitées par ordinateur dans presque toutes les disciplines. 2) Le fait qu'il existe des bases de données discrètes dans plusieurs disciplines. 3) La tendance vers la publication et la diffusion facilitées par l'utilisation de l'ordinateur, à vrai dire géré par ordinateur. 4) La tendance vers le transfert de données au lieu du transfert de documents. 5) Le rôle prépondérant de l'informatique de gestion à l'intérieur d'une bibliothèque pour contrôler les systèmes d'opération. La technologie appelle un constant effort de mise à jour. Les "Generalized Data Management Systems" fournissent la structure essentiele aux nouvelles découvertes concernant le traitement des données par les bibliothèques et les centres d'information. Un de ces systèmes, DATA-MAN,s'appliquant à une base de données à circulation interne dans la bibliothèque constitue l'exemple parfait de la simplicité et de la flexibilité et de l'utilité de cette approche. Une connaissance approfondie des principes de gestion des données est essentielle au bibliothécaire moderne. Des cours couvrant ces sujets doivent faire partie intégrante des programmes d'études des écoles de bibliothéconomie.

DATA - THE MISSING LINK ...

This afternoon I want to single out an area that I believe to be of great importance, that employs a technology well within the state-of-the-art, that is being widely used in industry and government and to a lesser degree in universities. The area is Data Management; more specifically, Data Management using Generalized Data Management Systems.

But before getting into Data Management and specifically Generalized Data Management Systems, I want to describe two recent experiences.

DATA - THE MISSING LINK ...

In 1968 I was Manager of Technical Information Services for the Producing Department of a major Canadian oil company. Our department consisted of three special libraries or information centers which served the literature needs of our administrative, technical and research staff. Another department, Technical Data Services, served the data requirements of the same user community.

Rather than try to differentiate between "information" and "data" from a theoretical standpoint, I shall give you a typical example which I hope will illustrate the operational distinction that existed at that time and place.

Suppose a geologist was interested in attempting to determine the depositional environment of a formation by measuring the salinity of its connate water. This geologist might well begin by asking Technical Information Services to find what internal reports, published papers, etc. were available on this subject. He might also go to Technical Data Services and have them run a search of the computerized well-data system for relevant data. Or indeed, he might do both.

Well, my feelings on this dual approach were precisely what I detect some of you are thinking: "Why not one-stop shopping?" Why not give this user the relevant documents which represent someone's synthesis and at the same time provide him with a data listing which represents someone's observations or measurements relevant to the phenomena under study.

I proposed the integration of TIS and TDS to management in much the same way as I am proposing it to you but as we all know organizations change slowly and integration was not to be in my time there. I resigned in 1969 and in 1972 integration of TIS and TDS did come about. Reports say that it is a satisfactory union and both departments are using the same Generalized Data Management System, Mark IV, for both document and data retrieval.

However, my problems have a way of following me around. In the Fall of 1971 I was Head, Department of Computer Services at the University of Calgary when Mr. Harry Campbell of Toronto Public Library called into my office. He had in his brief case a freshly printed compendium of machine-manipulable "hard data" bases in the Social Sciences in Canada. (Toronto Public Library

(1971)). He was looking for a home not only for his compendium but more to the point for the machine-manipulable "hard data" bases described within it or at least the subset of them that were actually or potentially relevant to U. of C.

I couldn't give Mr. Campbell much consolation. We at the computer center were not geared to be a library or information center and the library was not yet geared to handle this type of message or medium. He had an orphan on his hands, at least for the present.

However, these two experiences made me adopt the view that libraries individually or better still collectively, had better start gearing up to handle data in a big way. There are at least five compelling reasons for such a development program:

1. The rapid proliferation of bibliographic data bases in machine-manipulable form for both current awareness and retrospective search, e.g. Compendex.

2. The growth, both in quantity and quality, of "hard data" bases which contain raw (a partially cooked) data, e.g. Census Tapes, Social Science Data Bases.

3. The progress that is being made in data communications in terms of both effectiveness and efficiency over document transfer.

4. The trend toward computer-assisted, indeed computer-based, publication and dissemination. Carroll and Tague (1973).

5. The acute need for libraries to use data management systems to help plan, direct, organize and control their internal systems, e.g. Acquisitions, Circulation, etc.

Given these facts, how can we proceed?

The two major ways to develop the kind of data management power required are illustrated in Fig. 1.

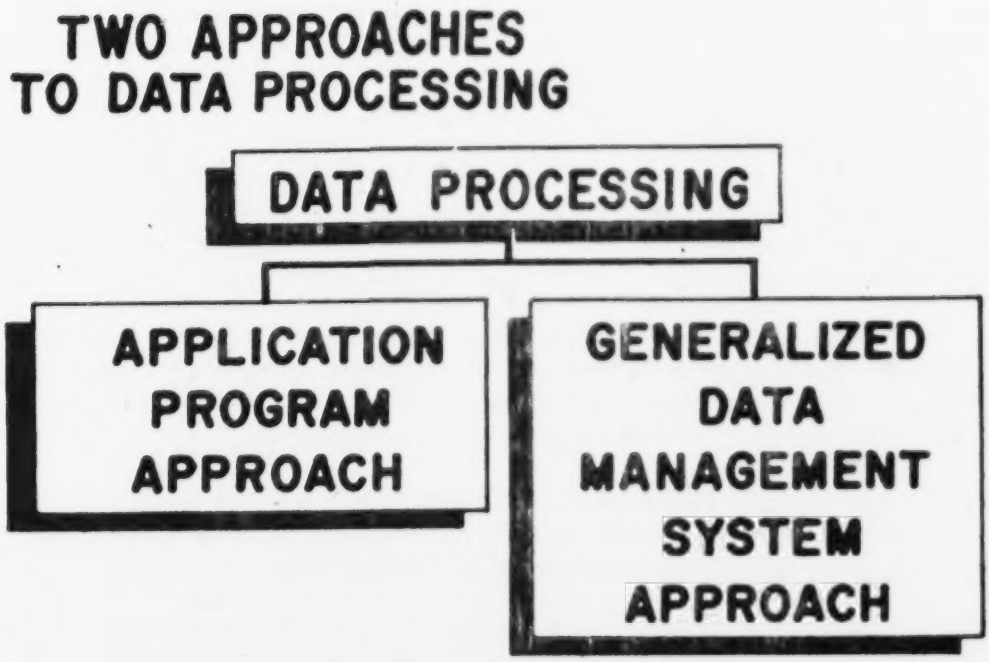


FIG. 1

The first approach is by Applications Programming. This is the conventional approach which involves well formulated objectives and rather rigid input and output formatting. A program, or series of programs, is written in a language such as COBOL to do a specific job. System changes are generally done in the form of patches until what you have looks more like an old bicycle tube than a system.

The second approach to the development of data management is to acquire a good Generalized Data Management System such as Mark IV, Assist, DATA-MAN, etc. from a reliable software manufacturer. Dolan, Dust, Woodcock (1970). You should make sure that the manufacturer is well established since even the best systems require constant maintenance and development.

To show you the power of Generalized Data Management Systems in terms of simplicity, flexibility and relevancy to library

operations, I shall demonstrate the application of a GDMS system, DATA-MAN, to the data management of an internal library system. DATAMAN (1973). It might well be possible to develop a substrate data bank to support all the operational systems in a library. However, for our modest example we have chosen a simple circulation data base.

The layout of a transaction as recorded in this circulation file is shown in Fig. 2.

LAYOUT OF CIRCULATION CARD

TRANS #	CALL NO			TITLE (40ch)	BORR #	BORR DATE			DATA ELEMENTS
	CLASS #	PUB YR	CPY #			YR	MO	DA	
									FORMAT
0 0 0 0 1	2 3 5 0 5	9 7	0 1	DATA PROCESSING FOR DECISION MAKING	0 0 0 2 6 8	7 3	0 3	1 7	EXAMPLE

FIG. 2

The top band shows the data elements or fields which make up the record, the middle bands show the format of the record, and the bottom band shows an actual example of the transaction data.

In order to enable DATA-MAN to process this file we must first describe the file to the system. Fig. 3 shows how this file definition would be coded.

DATA-MAN SYSTEM SPECIFICATIONS PAGE NO. 01 of 03

DEPARTMENT NO. JOB DESCRIPTION CIRCULATION FILE NO. 001

Line	DIRECTIVE	NAME	PARAMETER
9	12	24 130	36 41 46 51 56 61 66 71
010	FILE	CIRCULATION	
020		RECORD SIZE	68.
030		BLOCK SIZE	680.
040			
050	IDENTIFY	CIRC. RECORD	
060	DEFINE	TRANS NO	6,Z
070		CALL NO	10,A OVERLAY(01)
080		CLASS	5,Z
090		PUBL. YEAR	2,Z
100		COPY NO	2,Z
110		TITLE	40,A
120		BARCODE	6,Z
130		DATE	6,A OVERLAY
140		YR	2,Z
150		MO	2,Z
160		DA	2,Z
170	END		
180			
190		CIRCULATION FILE DEFINITION	
200		(DATA-MAN)	

FIG. 3

Once the file has been described and made available to the system we can process it. The same coding sheet that we used to define the file can now be used to write our processing instructions or program. Programming a GDMS system is much easier than programming in a programming language. To indicate the processing capabilities of the system, the directives which may be used in DATA-MAN are listed in Fig. 4.

SYSTEMS DIRECTIVES

AND	ELSE	LOOKUP	SELECT
AND NOT	END	MODIFY	SELECT IF
CALC	FILE	NEW PAGE	SELECT IF NOT
CALCULATE	FILEF	NO PRINT	SELECT NOT
COMMON	FORMSCODE	NO SELECT	SET MARGIN
COPY	GO TO	OR	SET TAB
CREATE	HEADING	ORNOT	SORT
DEFINE	IDENTIFY	PATTERN	SUBHEADING
DEFINE SUM	IF	PRINT	SUM
DELETE	IFNOT	PRINT ZEROS	SUMMARIZE
DESCRIBE	IN ANY CASE	PROCESS	TABLE
DETAIL	LINE	RENUMBER	TOTAL
EDIT	LINEF	REPORT	TURN ON
EDIT WIDTH	LINE COUNT	RETRIEVE	TURN OFF

FIG. 4

DATA - THE MISSING LINK ...

Getting back to our application of DATA-MAN to the circulation file suppose that we were interested in finding out how much activity there was in a certain class range in our library. Let this class internal be 95400-95499 inclusive, which incidentally stands for "Systems Theory" in our small Library-School Library.

Fig. 5 shows how we would code up this request, again using the same type of form.

DATA-MAN SYSTEM SPECIFICATIONS PAGE NO 01 OF 03

DEPARTMENT NO. 1 JOB DESCRIPTION ...

NO 01

LINE	DIRECTIVE	NAME	PARAMETERS
0	12	24 30	36 41 46 51 56 61 66 71
010	RETRIEVE	CIRCULATION	
020			
030	SELECT IF	CLASS	95400-95499
040			
050	SORT	CLASS	
060			
070			
080			
090			
100			
110			
120			
130			
140			
150			
160			
170			
180			
190			
200			

CIRCULATION FILE RETRIEVAL (DATA-MAN)

FIG. 5

Finally we need to specify that we wish to generate a report and describe the format for this report. Fig. 6 shows the coded specifications for report generation.

DATA - THE MISSING LINK ...

DATA-MAN SYSTEM SPECIFICATIONS PAGE NO 03 of 03

DEPARTMENT NO. 1 JOB DESCRIPTION ...CIRCULATION FILE NO CIRC

LINE	DIRECTIVE	NAME	PARAMETER
9	12	20 30	36 41 46 51 56 61 66 71
0,10	REPORT		
0,20			
0,30	HEADING		'SPACE 1'
0,40			'DOCUMENTS CIRCULATING IN CLASS'
0,50			'INTERVAL 95400-95466 MARCH 1973'
0,60			
0,70		$DATE	
0,80		$PAGE	
0,90			
1,00	DETAIL		
1,10		TRANS. NO.	'TRANSACTION', 'NUMBER'
1,20		CLASS	'CLASS'
1,30		PUBL YEAR	'PUB', 'NO'
1,40		COPY NO	'COPY', 'NO'
1,50		TITLE	'TITLE'
1,60		BORROWER	'BORROWER'
1,70		YR	'YEAR'
1,80		MO	'MONTH'
1,90		DA	'DAY'
2,00			
2,10	TOTAL	$RECORD	
2,20		$COLUMN	
2,30	END		
		CIRCULATION FILE REPORT FORMATTING (DATA-MAN)	

FIG. 6

The resulting report is shown in Fig. 7

DOCUMENTS CIRCULATING IN CLASS INTERVAL 95400-95466 MARCH 1973

TRANSACTION NUMBER	CLASS	PUB NO	COPY NO	TITLE	BORROWER	YEAR	MONTH	DAY
001063	95405	968	01	MANAGEMENT SYSTEMS	101928	73	03	09
001129	95475	970	14	LIBRARY SYSTEMS ANALYSIS GUIDELINES	498525	73	03	14
002071	95475	972	01	OPERATIONS RESEARCH IMPLICATIONS FOR LIB	806930	73	03	29
001236	95490	969	01	SYSTEMS ANALYSIS FOR BUSINESS DATA PROCE	498525	73	03	16
001349	95490	969	02	SYSTEMS ANALYSIS FOR BUSINESS DATA PROCE	000264	73	03	20

RECORDS PROCESSED 627
RECORDS SELECTED 5

CIRCULATION FILE REPORT
(DATA-MAN)

FIG. 7

DATA - THE MISSING LINK ...

Thus you have been exposed to the three main phases of data management using a Generalized Data Management System. In review, these phases are: File Definition, File Processing and Report Generation.

You can see that these systems are simple to use and can be applied to bibliographic data bases, "hard data" bases of subject areas or administrative data bases internal to the library.

GDMS Systems represent a significant development for libraries and information centers. The study of Data Management principles, practices and systems can no longer be a missing link; it must be forged into its rightful place in the evolution of Library School Curricula.

REFERENCES

DATAMAN LTD., 1973, DATA-MAN File Management Systems (Documentation).

DOLAN, F.T., DUST, N.H., and WOODCOCK, J.F., 1970, Evaluation of Generalized Data Management Systems. The University of Calgary.

TAGUE, J. and CARROLL, J., 1973, A Computer-Based Alternative For Scientific and Technical Publishing. Typed manuscript. Financed by NRC, Grant No. A7132.

TORONTO PUBLIC LIBRARY, 1971, Inventory of Social Science Quantitative Data Sources In Canada. Prepared under a contract with the Social Science Research Council of Canada.

TEACHER EDUCATION IN INFORMATION SYSTEMS UTILIZATION (PREPARATION DES PROFESSEURS POUR L'UTILIZATION DES *SYSTEMES D'INFORMATION*)

Henry James Abraham Goodman
Faculty of Education, University of Calgary
Calgary, Alberta T2N 1N4

ABSTRACT

The University of Calgary Faculty of Education is attempting to enable future teachers to cope with the information revolution-explosion. Much of the "Communication in the School" course is devoted to those aspects of information science related to the teaching-learning process, including: information communication theory and models; the production and use of non-print and print resources; and biliomediagraphic and computer assisted instruction systems utilization training. (La Faculté d'Education de l'Université de Calgary essaie de préparer les instituteurs de l'avenir à faire face à la révolution et à l'explosion dans le domaine de l'information. Une bonne partie du cours "Communication dans les Ecoles" est consacré à l'étude et à l'enseignement de divers aspects de la science d'information dont la théorie et certains modèles d'information-communication; la production et l'usage de resources imprimées et non-imprimées; l'utilization de systèmes pour l'enseignement assisté par ordinateur et des systèmes pour l'enseignement assisté de bibliographies et de médiagraphies.)

TEACHER USE OF INFORMATION SYSTEMS

Just as good child health care is often said to begin in the pre-natal period, so too, the appropriate point for the proper inception of the educator is the moment when he begins his professional education. And what better nurture is there for the embryonic teacher than that which can not only assist in sustaining his growth during the gestation period, but can also provide him with a substantial basis for continued intellectual growth throughout his professional career - the ingestion of a modicum of information concerning how to cope with one crucial facet of the information revolution-explosion, namely, information about, and some skill in, information systems utilization?

The Faculty of Education at the University of Calgary has instituded an academic year-long course for freshmen entitled "Communication in the School" comprised of four quarter or mini courses. In toto, well over a quarter of the entire course consists of content devoted to aspects of information science as they relate to the teaching-learning process. The general introductory sessions, for example, deal with such topics as information and communication theory and models. It is the "Media" quarter or "element" of the course, involving three hours per week for twelve weeks, which is the particular concern of this paper. Here it is that the principal effort is made to lay the foundation whereby future teachers will ultimately be converted into professionals with expertise in the use of manual and automated information storage and retrieval systems and computer assisted instruction systems as they pertain to educational communication. Herein are treated the various forms of enhancement of teacher-learner communication in which the media serve as producers, storers, retrievers, transmitters, or displayer-presenters of information.

It is particularly noteworthy that the quarter course under discussion here is officially designated as the "media" element. This designation was deliberately selected to be indicative of the fact that the substantive content to be treated includes not only the print media but also, and even more emphatically, the "audiovisual" media including such newer electronic mass media as cable television. This deliberate emphasis on the audiovisual domain is not given merely to redress the imbalance of treatment vis-à-vis print and non-print media commonly found in information and communication science studies, particularly information retrieval studies. It is based upon the notion, that for a variety of reasons, such materials are, by and in the large, under-used and often misused in formal educational processes, despite the fact that the number and variety of these materials designed for school use appear to be growing exponentially. The fact that this apparently contradictory situation persists, concurrently with the strong and pervasive audiovisual media orientation, if not saturation, of the current generation of students, prospective teachers, and experienced teachers in "out-of-school" situations, renders the task of implementing a change in these matters all the more imperative.

TEACHER USE OF INFORMATION SYSTEMS

Among the more significant factors resulting in the ineffective and/or infrequent use of audiovisual resources in education is the widespread lack of adequate information about these resources, both at the classroom level and the school instructional materials or resource center (library) level. Doubtless, too, the situation resulting from the relative paucity of available secondary-reference materials dealing with audiovisual materials is exacerbated by the lack of available tools for the selection of these non-print resources, or even agreed upon criteria for their selection. There is all too often a resultant dependence upon unreliable producers' or publishers' catalogs for media resource information. There thus exists a distinct "knowledge gap" about audiovisual sources of knowledge.

Attempts to meet these "mediagraphic control" needs have ranged from simple linear files in libraries to coordinate indexing systems of inverted files, and, recently, to automated systems based on computer programs operating on magnetic tape or disk storage of data bases. Such attempts, quite difficult at any level of sophistication, even in the case of the print media, are rendered exceedingly more so, and more costly too, in the case of multi-media and/or multi-dimensional resource formats. The latter strictures apply equally to the classification and to the storage and retrieval aspects of these media.* Moreover, the existing small number of successful automated attempts on the part of information services to meet these needs are still relatively little known to many teachers. Consequently, the media element is specifically designed to provide the student with a "head start" in terms of knowledgeability of secondary reference materials and information systems dealing with the increasingly important audiovisual media useful in education. It is first intended to inform him of the printed means whereby he can discover and discriminatingly select from the existing audiovisual media those items suited to his students and to his learning objectives: Knowing how to make use of citation indexes; being aware of special reviews, and of various computerized services, such as SDI (selective dissemination of information) services; being aware of alerting services; and knowing the extent to which these services may be obtained

---

*The Task Force on Computerized Cataloging and Booking of Educational Media Report (Standards for Cataloging, Coding, and Scheduling Educational Media, Washington, D.C., Dept. of Audiovisual Instruction (NEA), 1968) provided the basis for film library storage and retrieval of information based on titles, subjects, and specific content codes, all of which are applicable to content which can be codified in terms of the words used in verbal communication. But where the content to be retrieved consists of the materials of non-verbal communication, of as yet undefined terms, of rapidly changing dynamic content, of a multi-dimensional nature, such as computerized dynamic holographic visual presentations, the task of codification and of selective retrieval still remains to be systematically attacked.

automatically and inexpensively would presumably facilitate the critical selection of print and non-print materials used in the classroom. The element similarly seeks to prepare the student for the task of using such non-print reference materials and associated equipment as are currently available, for example, microform catalogs of audiovisual materials, and microform readers, again as tools to selectively retrieve information from both print and audiovisual sources.

Hopefully, then, the future teacher will be at least partially prepared not merely for the end of the "search among the multiple card catalogs and producers' or publishers' book catalogs and ye shall find" era but more importantly, for the nascent era of automated information technology, wherein he will doubtless live and work. Hopefully, too, the groundwork will have been laid for both a willingness and a capability on his part to render assistance in the finding of solutions, immediate and more long range, to the information retrieval problems with which he, as a teacher, will soon doubtless be confronted. Efforts are, furthermore, exerted to have the prospective teacher become aware of the probability of the eventual existence and availability to him, his students, and the public at large, at home, in the school, and at places of business, of computer-based central data information storage facilities as well as information retrieval networks. The learning of the operation of teletype computer assisted instruction terminals is therefore regarded as partial preparation for the anticipated expertise required in the functioning of the multi-media terminals by means of which the local, regional, national, and international retrieval networks will doubtless be accessed.

Other basic operational skills necessary for the effective use of the devices and tools of modern educational technology, not already noted above, are also imparted. Flowing from such training, it is hoped, will be the subsequent assured acceptance and efficacious use of both currently available and yet-to-be developed viable technological aids.

At the outset of the media element its objectives are made clear in a formal statement, a copy of which is provided to each student. A portion of that statement is summarized as follows:

At the conclusion of the media element the students are expected to be able to:

1) describe the "novel" aspects of the newer media which distinguish them from information storers, retrievers, organizers, and carriers such as books and chalkboards;
2) discuss in an essay the potential repercussions of the acceptance of computer assisted instruction as a form of education, in terms of its effects on teachers, learners, and society;
3) describe how learning materials or resources may be used

effectively by teachers to enhance student learning; and
4) describe the distinctive features of several indexes and other secondary sources of media materials, print and non-print, including those which involve microforms and/or computer tapes.

The attempted actual attainment of these goals involves the student as a participant rather than as an audience member in numerous very specific activities, including the completion of certain assignments and projects. These experiences are all designed to provide opportunities, as extensive as possible within the prevailing time limits, to practice and to analyze information-communication skills in situations which are themselves as practical as is physically feasible.

Many of the information retrieval-type learning experiences arranged for the prospective teacher involve activities which have their locale within the Instructional Materials (or Resources) Center of the Curriculum and Instruction Department. These activities are preceded by an audiovisual tape-slide presentation-orientation session at which printed hand-outs dealing with the Center and a related major information retrieval assignment are distributed. Essentially, the assignment deals not so much with the use of the facilities and tools commonly found in information centers or libraries, such as standard encyclopedias and periodical indexes, but rather, with special types of secondary information sources which bring together print and non-print teaching-learning curricular resource materials under subject or media format, or other rubric, including indexes, directories, card or book catalogs, abstracts, thesauri, descriptor lists, mediagraphies, and media review periodicals.

The student is expected to engage in individual and independent effort and to use all of his own resources and ingenuity before approaching the Center personnel for assistance. He must first select a specific topic regularly treated in his intended subject-matter teaching field and grade level. He then discovers how to make use of appropriate secondary sources, so as to obtain the relevant mediagraphic and/or bibliographic details concerning a given number of recent teaching and learning resource materials dealing with the selected topic and grade level.

The preceding assignment questions are so phrased as to involve the finding of answers to Taube's (1956) three now classic types of questions which a searcher might want to ask when seeking information from an information system:

1) Index type question: "What 'documents' are available on the specific topic in question?";
2) Content question: "What subjects (i.e., topics and sub-topics) are covered in the extant 'documents'?"; and
3) Association question: "What other topics are relevant to the precise

subject being searched?".

The would-be teacher here gains familiarity with such reference sources as the seven volume Westinghouse Learning Directory, a work based on a computerized information bank. It provides curriculum topic or unit-related information for immediate reference by the learner and/or teacher. It is at this point, too, that he is likely to first encounter the NICEM (acronum for National Information Center for Educational Media) Catalogs, each volume of which treats the resource material available in a particular non-print media format, such as the filmstrip format. The fact that neither of these works provides clues as to the quality of the items listed therein is one that it is intended the student will note. He must then find one appropriate non-print and one print item among the holdings of the Materials Center which are not in circulation and must preview and evaluate these two items. To that end his attention is drawn to particular articles in information sources which not merely list but also critically review the resource materials with which they deal. Hopefully, some of the evaluative criteria, implicit and explicit, found therein, "rub off" on the teacher trainee. As a result, he is expected to be able to answer such queries as: "Would the students of the given grade level understand the content?"; "Would they be likely to find it interesting?"; "Does the content appear to be accurate?"; "Is it well-produced?". He is additionally required to adduce evidence or reasons for his responses to these questions.

Special attention is given to the ERIC system of the U.S. Office of Education as an example of a particularly useful educational information system. Knowledgeability concerning ERIC offers the teacher-aspirant an opportunity to assess his own professional identification level. ERIC can do so because it provides the members of the teaching profession with one of the basic means for the fulfilment of their desires as professional people - to become involved in a process of continuous learning and to "keep-up" with the advancements in their field. ERIC at last makes available at moderate cost and with dispatch professional services in education matching those previously available primarily in scientific fields.

A hand-out entitled "What ERIC is about, what it does, and how it does it" is distributed to each student. Included in it is a note indicating that the main Library of the University has a microfiche copy of the complete text of virtually every "document" listed in the ERIC system, and that the student also has access, at minimal cost, to a computerized magnetic tape information retrieval service which utilizes ERIC "documents" as its data base. As the latter system is not as yet "on line", the student is advised that the system will be accessible to him only through human intermediaries in the form of the University's Information Services Division and through use of the specific search terms employed in the ERIC system.

TEACHER USE OF INFORMATION SYSTEMS

To prepare the future teacher for the possible advent of a world-wide, user-oriented, single source information system such as that already in embryonic existence in the form of UNISIST (the United Nations Scientific Information System) and/or an encyclopedic educational organism such as the projected Wellsian "World Brain", to which he, both as a teacher and as a citizen, would presumably have access, and, to have him also give serious consideration as to whether he should either lend his support to or engender opposition to such proposals, a major specific subjective open-ended question is included in the assignment. The question is worded as follows: "What potential for the classroom teacher do you see in ERIC as a model of an even broader and more inclusive educational information system?".

Apart from the Materials Center assignment, other forms of practical information retrieval experience are provided within the media element. In the case of mediated information sources, primary or secondary, user knowledge of the operation of currently widely used "hardware" items, such as film, filmstrip, and slide projectors, microform readers, audio and video tape recorders, record players, and teletype computer terminals, is imperative. Each student is accordingly provided with both a demonstration of and a hands-on experience in the operation of these devices. In addition, training in the oeration of random access, remotely controlled, and other more sophisticated equipment, such as multi-media programmers, is provided on an optional basis.

The media element as well as the other elements of the course are all undergoing more or less continuous revision on the basis of regular faculty and student evaluations. Close collaboration of faculty colleagues with prime responsibilities in the educational aspects of information science - the instructors in audiovisual media courses and those in school librarianship, together with the professional staff of the Materials Center - will doubtless result in further major modifications in the media element. Several innovations are either immanent or are being contemplated for future implementation. It is hoped, for example, that during the 1973-74 session the University of Calgary will be one of the first institutions to install the very sophisticated and very comprehensive on-line IBM information retrieval system "STAIRS" (an acronym for Storage and Information Retrieval System) and that our teacher trainees will be among the first to have an opportunity to familiarize themselves with it as potential users. Beyond that, it would seem desirable to provide our students with yet another form of training. Having, in due course, acquired the basics of media "software" production, it would be well if they could also learn how to input into the retrieval system information regarding their own newly-produced resources. By adding their informed efforts to those of their prospective colleagues, teacher education students will hopefully help themselves and their future profession to cope with the information revolution-explosion.

REFERENCE

TAUBE, M., 1956, Studies in Coordinate Indexing, v. 5, Belhesda, Md., Documentation Inc.

L'EXPERIENCE DE L'UNIVERSITE LAVAL DANS L'UTILISATION DES BANQUES D'INFORMATION (LAVAL UNIVERSITY'S EXPERIENCE WITH DATA BASES)

R. Francoeur, E. Loranger, A. Gamache
Centre de documentation de la Bibliothèque
Université Laval, Québec G1K 7P4

RESUME

L'équipe Gesydams du Centre de documentation de la Bibliothèque de l'Université Laval a pour but de faciliter aux usagers l'accès aux banques d'information spécialisées. Elle établit un lien direct entre les usagers et les systèmes. Son rôle consiste à dépister, signaler et exploiter les services de dissémination sélective de l'information. En avril 1973, l'équipe gère plus de 145 profils courants et rétrospectifs. (The aim of the Gesydams team at the Documentation Center of Laval University's Library is to facilitate users' access to specialized data bases. Gesydams establishes direct links between users and systems. Gesydams' role is to screen, to report on and to operate, generally speaking, selective dissemination of information (SDI) services. At present (April 1973) the team is managing over 145 current and retrospective interest profiles).

GESYDAMS

INTRODUCTION

Depuis 1967, le Centre de documentation de la Bibliothèque de l'Université Laval offre aux chercheurs un service de dissémination sélective de l'information (DSI). Afin de concilier les exigences de la multidisciplinarité du milieu universitaire et les contraintes imposées par nos ressources financiaires limitées, nous avons orienté, dès le début, ce service DSI principalement sur l'exploitation des banques d'information existantes où qu'elles soient.

En 1968, la création de CAN/SDI permit d'élargir l'éventail des services documentaires du Centre et d'assurer aux chercheurs une signalisation périodique à un tarif rai-

sonnable. L'accueil réservé à ce nouveau service de signalisation fut encourageant et dès 1971, l'équipe du DSI fut désignée par le sigle GESYDAMS.

L'EQUIPE GESYDAMS

Le sigle GESYDAMS signifie: Groupe d'exploitation des systèmes documentaires automatiques. L'équipe est un groupe "Search Editor" dont les objectifs sont: promouvoir l'utilisation et faciliter l'exploitation des banques d'information.

FONCTIONNEMENT

Les principales fonctions de l'équipe GESYDAMS sont: le dépistage et l'analyse des nouvelles banques d'information, la promotion des services de signalisation, l'élaboration des profils d'intérêt, les communications avec les systèmes, l'analyse statistique des profils et leurs modifications. L'annexe A illustre le cheminement des opérations qu'effectue l'équipe GESYDAMS.

AVANTAGES DE CETTE FORMULE

L'équipe offre les avantages suivants:

Pour l'usager

1) Maintient un inventaire à jour des banques d'information
2) Conseille l'usager pour l'élaboration de son profil d'intérêt et le choix des banques d'information pertinentes à son domaine de recherche
3) Effectue l'analyse statistique du profil d'intérêt afin d'améliorer la pertinence des références signalées
4) Permet d'effectuer rapidement les modifications du profil d'intérêt.

Pour le système

1) Effectue les tests nécessaires (en soumettant des profils d'intérêt expérimentaux) permettant l'évaluation pragmatique des nouvelles banques d'information
2) Accélère le traitement des profils en élaborant plus rapidement une meilleurs stratégie de recherche
3) Favorise l'exploitation des banques d'information grâce à son programme d'information multi-media aux usagers.

DIFFICULTES RENCONTREES

Parmi les difficultés rencontrées, il faut aussi signaler:

1) La résistance des usagers, en particulier des chercheurs, à déléguer leur recherche documentaire à une tierce personne. Nous avons remarqué cependant qu'un profil expérimental permet de surmonter cette difficulté

2) Le scepticisme des chercheurs prétendant être informés suffisamment au moyen de contacts personnels. L'étendue de la couverture des banques d'information spécialisées, la variété des éléments de dépistage et la rapidité du service offert suffisent à aplanir les doutes persistants
3) L'inquiétude de certains chercheurs à être submergé par une documentation trop abondante. Cette crainte disparaît devant l'assurance de la disponibilité de GESYDAMS pour effectuer les modifications du profil.
4) Les lacunes de certaines banques d'information encyclopédique dans des secteurs spécialisés de la recherche. C'est alors que GESYDAMS suggère l'utilisation de banques d'information plus spécialisées.

ANALYSE STATISTIQUE

Lorsqu'un usager est insatisfait des références obtenues, il peut obtenir rapidement une analyse statistique de son profil d'intérêt. Il doit pour ceci, pondérer un échantillon de références reçues et les transmettre à GESYDAMS.

Les références sont alors regroupées par éléments de repérage et par banques d'information. La programmerie STAT permet d'évaluer les paramètres suivants:

Pour chaque descripteur i, simple ou coordonné
P_i : facteur de pertinence
I_i : facteur d'imprécision
E_i : facteur d'efficacité
BR_i : facteur de bruit

Pour le profil
EG = facteur d'efficacité
IG = facteur d'imprécision

$$P_i = \frac{B_i}{D_i} \qquad I_i = \frac{C_i}{D_i} \qquad E_i = \frac{R_i}{\sum_{i=1}^{i=n} R_i}$$

$$BR_i = \frac{N_i}{\sum_{i=1}^{i=n} N_i} \qquad EG = \frac{\sum_{i=1}^{i=n} B_i}{\sum_{i=1}^{i=n} D_i} \qquad IG = \frac{\sum_{i=1}^{i=n} C_i}{\sum_{i=1}^{i=n} D_i}$$

où B_i = nombre de références pertinentes signalées par le descripteur i
C_i = nombre de références imprécises signalées par le descripteur i
D_i = nombre total de références signalées par le descripteur i
N_i = nombre de références non-pertinentes signalées par le descripteur i

L'annexe B présente en première partie un exemple d'un profil d'intérêt à analyser.

La deuxième partie de cette annexe renferme les résultats de l'analyse statistique effectuée pour ce profil.

On remarque alors, que le rendement du profil se situe entre 23.88% et 36.56%.

Le profil signale donc une référence pertinente sur trois. En examinant les résultats des analyses statistiques pour chacun des descripteurs simples ou coordonnés, on remarque que deux de ces descripteurs donnent plus de 62% du facteur de bruit global. Ces descripteurs sont: DEVELOP*+CHILD* et FOOD*+HABIT*. Il est cependant impossible de songer à retrancher complètement ces deux descripteurs coordonnés puisque le facteur d'efficacité de ces deux descripteurs coordonnés représentent au-delà de 28% du facteur d'efficacité globale. Le facteur de pertinence de chacun de ces descripteurs coordonnés, confirme cette décision. En effet, on remarque que 12% des références repérées par chacun de ces descripteurs ont été jugées pertinentes.

En examinant les références dépistées par ces descripteurs, on remarque que le descripteur EDUCAT* apparaît sur la plupart des références non pertinentes repérées par le descripteur DEVELOP*+CHILD*. Pour le descripteur FOOD*+HABIT*, on remarque qu'en limitant la recherche aux descripteurs *MAN ou *MEN ou CHILD*, on pourrait ainsi éliminer la plupart des références non pertinentes.

La troisième partie de l'annexe B montre le profil d'intérêt modifié à l'aide des résultats de l'analyse statistique.

STATISTIQUES D'UTILISATION DU SERVICE GESYDAMS

En avril 1973, GESYDAMS dessert environ 12% des chercheurs de l'Université Laval. Ceci exige la gestion de plus de 90 profils courants et de plus de 55 profils rétrospectifs.

GESYDAMS

Le tableau suivant donne la liste des banques d'information ainsi que le nombre de profils rétrospectifs et courants.

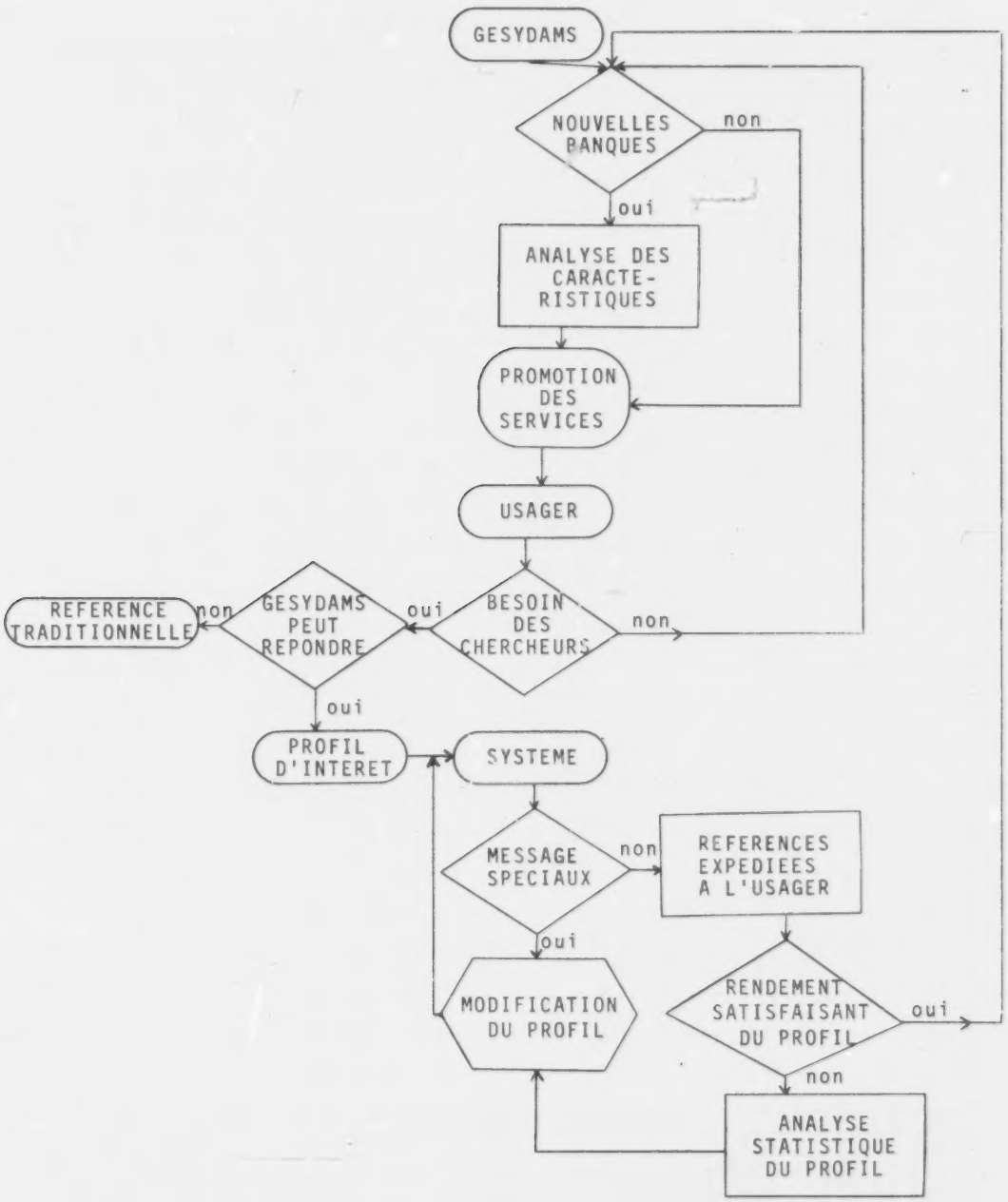
Courants	
ISI	35
Geo-Ref	4
GRA	2
Compendex	2
Inspec	6
Chemical Abstract	23
Medlars	8
Chemical Title	2
Biological Abstract	9
CAIN File	1
ERIC	1
Rétrospectifs	
ERIC	3
Medlars	28
Medline	1
Excerpta Medica	2
Compendex	2
Vibanque	23

CONCLUSION

L'accueil réservé par les chercheurs de Laval à la dissémination sélective de l'information a été amplement suffisant pour nous convaincre de continuer à offrir ce service. Le rayonnement de GESYDAMS a débordé les cadres de l'Université Laval. Plusieurs centres de recherche du gouvernement et des universités du Québec ont sollicité et obtenu ses services. Cette activité régionale du Centre de documentation s'inscrit dans le cadre de l'établissement d'un réseau national de centres régionaux DSI.

ANNEXE A

SCHEMA DU FONCTIONNEMENT DE GESYDAMS



ANNEXE B

1) PROFIL A ANALYSER

Université Laval
BIBLIOTHÈQUE - CENTRE DE DOCUMENTATION

GESYDAMS

QUEST NO	et ou	DESCRIPTEURS	et ou	DESCRIPTEURS	et ou	DESCRIPTEURS
1	A	VEGETARIAN				
	B	NIBBLING				
2	C	DEVELOP*	D	*NUTRITION*		
			E	CHILD*		
			F	CURVE*		
3	G	DIET*	I	PATTERN*		
	H	*FEEDING*	J	CONSUM*		
			K	HEALTH*		
			L	POOR*		
			M	HABIT*		
			N	DISADVANTAG*		
4	O	FOOD*	P	FAD*		
			J	CONSUM*		
			M	HABIT*		
			K	HEALTH*		
			L	POOR*		
5	Q	FREQUENC*	R	MEAL*		

NE	B	NR	EXPRESSIONS DE RECHERCHE
01	I	99	A/B
02	I	99	C&(D/E/F)
03	I	99	(G/H)&(I/J-N)
04	I	99	O&(P/J-M)
05	I	99	Q&R

2) ANALYSE STATISTIQUE DU PROFIL

① DESCRIPTEURS	② B	③ C	④ D	⑤ P %	⑥ I %	⑦ E %	⑧ BR %
VEGETARIAN	6	0	6	100	0	18.75	0
NIBBLING	1	0	2	50	0	3.12	1.17
DEVELOP*+CHILD*	7	7	55	12.72	12.72	21.87	48.23
DEVELOP*+*NUTRITION*	4	3	10	40	30	12.05	3.52
DEVELOP*+CURVE*	0	0	5	0	0	0	5.88
DIET*+PATTERN*	1	0	6	16.66	0	3.12	5.88
DIET*+CONSUM*	0	1	3	0	33.33	0	2.35
DIET*+HEALTH*	2	1	9	22.22	11.11	6.25	7.05
DIET*+POOR	0	0	2	0	0	0	2.35
DIET*+HABIT*	0	1	1	0	100	0	0
FEEDING+DISADVANTAG*	1	0	1	100	0	3.12	0
FEEDING+CONSUM*	0	0	1	0	0	0	1.17
FEEDING+HABIT*	0	0	1	0	0	0	1.17
FEEDING+HEALTH*	0	0	1	0	0	0	1.17
FOOD*+CONSUM*	4	1	11	36.36	9.09	12.05	7.05
FOOD*+FAD*	1	0	1	100	0	3.12	0
FOOD*+HABIT*	2	3	17	11.76	17.64	6.25	14.11
FOOD*+HEALTH*	0	0	3	0	0	0	3.52
FOOD*+POOR*	0	0	1	0	0	0	1.17
FREQUENC*+MEAL*	3	0	3	100	0	9.37	0

⑨ EG = 23.88% ⑩ IG = 12.68%

LEGENDE

1 : Liste des descripteurs simples ou coordonnés
2 : Nombre de références pertinentes
3 : Nombre de références imprécises
4 : Nombre total de références
5 : Facteur de pertinence
6 : Facteur d'imprécision
7 : Facteur d'efficacité
8 : Facteur de bruit
9 : Facteur d'efficacité du profil
10 : Facteur d'imprécision du profil

3) PROFIL MODIFIE

Université Laval
BIBLIOTHÈQUE - CENTRE DE DOCUMENTATION
GESYDAMS

QUEST NO	et ou	DESCRIPTEURS	et ou	DESCRIPTEURS	et ou	DESCRIPTEURS
1	A	VEGETARIAN				
	B	NIBBLING				
2	C	DEVELOP*	D	*NUTRITION*		
			E	CHILD*	S	¬ EDUCAT*
			F	CURVE*		
3	G	DIET*	I	PATTERN*		
	H	*FEEDING*	J	CONSUM*		
			K	HEALTH*		
			L	POOR*		
			M	HABIT*		
			N	DISADVANTAG*		
4	O	FOOD*	P	FAD*		
			J	CONSUM*		
			K	HEALTH*		
			L	POOR*		
			M	HABIT*	T	*MAN
					U	*MEN
					E	CHILD*
5	Q	FREQUENC*	R	MEAL*		

NE	B	NR	EXPRESSIONS DE RECHERCHE
01	I	99	A/B
02	I	99	C&(D/F/E¬S)
03	I	99	(G/H)&(I/J-N)
04	I	99	O&(P/J-L/M&(T/U/E))
05	I	99	Q&R

COMMUNICATIONS AND THE WIRED CITY

L. Durr
Department of Communications
Ottawa, Ontario K1A 0C8

ABSTRACT

The attention which is now being given to the future communications networks in Canada in such areas as computer/communication and broadband wired city applications does not in the least exaggerate their importance. The long term impact of communications on such a scale truly challenges the imagination and provokes the far sighted into predictions of the wonderful -- or terrible -- consequences. It is important that these consequences be examined not only in terms of the contribution to improved communications but also in terms of the wider impacts upon society and the environment.

This paper describes some of the programs sponsored by the Department of Communications with the intention of both encouraging the production of these systems and the evaluation on a multidisciplinary plane.

MODULARITY, MICROFORM & MICROCOMPUTERS

J. W. Cherry & E. V. Goodman
Exxon Corporation
New York U.S.

ABSTRACT

When after over fifteen years' experience in developing and operating Information Retrieval systems Exxon Corporation developed its fourth generation system, LIIMS, it built a high degree of modularity into the system. One of the major modules, the Index Management Module, can be regarded as consisting of up to seventeen sub-modules. The user can select the sub-module or sub-modules which best meet his specific needs and vary this selection from task to task without penalty or programming demands.

This modularity has made it possible for Exxon to develop a Microform Information Transfer Center and exploit the new microcomputers with minimal systems effort. Searches of COM-produced indexes in the MITC lead to the recovery of the full-text of appropriate documents. The use of microcomputers has speeded up the indexing process, eliminated key punch and greatly facilitated editing, thesaurus construction and on-line printout. Equally exciting, it has simplified access to external information banks.

MODULARITY, MICROFORM & MICROCOMPUTERS

Exxon's Logically Integrated Information Management System (LIIMS) owes a great deal to the experience Ed and I gained developing and operating three Information Retrieval Systems--SIS I, SIS II, and SIS III, which originated in Canada. When we were assigned the task of developing LIIMS, we were in the fortunate and possibly unique position, for people working in the information field, of being invited to develop a fourth generation system after over twelve year's experience of working with computer-orientated information systems.

All of this experience suggested two things. First, that our fourth generation system should be user-oriented; second, that it should be highly modular in construction.

Let me address myself to the first of these premises. SIS I in its day was a fairly advanced system. We could make it perform a number of useful tasks. However, in Imperial's Technical Information Services Department, we had to maintain a "resident specialist" who made out complicated run-sheets for every piece of output we requested from the system. During the "up-date and run" period of our routine operation, he was in constant communication and often, conflict with a "resident specialist" in our Systems and Computer Services Department. SIS II and SIS III, to the best of our knowledge, did little to remove this barrier to system exploitation.

In the Index Management Module of LIIMS, we have removed this barrier completely. The user of the system, electing to choose between two mini-program decks, can run a pilot program, or have the computer produce any index to his specific requirements, directly. We can train him to use these mini-decks in half a day, and then he is completely independent.

We developed this approach before we began to use our micro-computer. I should emphasize, in fairness to Ed who designed the front end programs that made it possible, that simplifying the user/system interface calls for some very sophisticated programming, and that the introduction of these programs was a pretty traumatic "happening" for the client-run coordinators whose role in life up to that time was to act as a human interface between the user and the system. At the same time, I should point out that his approach is now being universally adopted by our Maths, Systems, and Computer Department, and that it greatly facilitated our move to the micro-computer.

Now let me address myself to the second premise - a modern IR system must be modular in design.

If anything confirmed our early suspicions about modularity, it was working with SIS II. SIS II, as I hope most of you know, was a

co-operative effort of Imperial, the Canadian Geological Survey, the U. of Saskatchewan, and the U. of Calgary. The programs were written by the U. of Saskatchewan and have always been regarded as extremely effective programs. The trouble with SIS II as a system however, is that if you want to make any changes to its routines or its output, you almost have to take it to pieces, modify it, and then put it back together again. If we learned anything from SIS II, it was that you had to be able to modify without dismembering - if you wanted a viable system.

At the same time as we were beginning to realize this basic fact, our ideas about "Information Banks" which we had first discussed at a meeting at the U. of Manitoba in September 1967, were beginning to crystallize and force us into modularity.

We define an "Information Bank" as a discrete collection of information maintained, organized and accessed with the assistance of a computer, pragmatically separated from other collections of information maintained, organized and accessed by the system because of differences in discipline, vocabulary, usage or even client preference.

If you think about that for a moment, you will realize we are inferring that even input to an IR system should be modular.

With all this in mind, we built seventeen modules into the Index Management Module of LIIMS.

The Index Management Module itself may be regarded as one of five major modules of LIIMS. Figure 1 shows the IMM modules in summary form.

INDEX MANAGEMENT MODULES

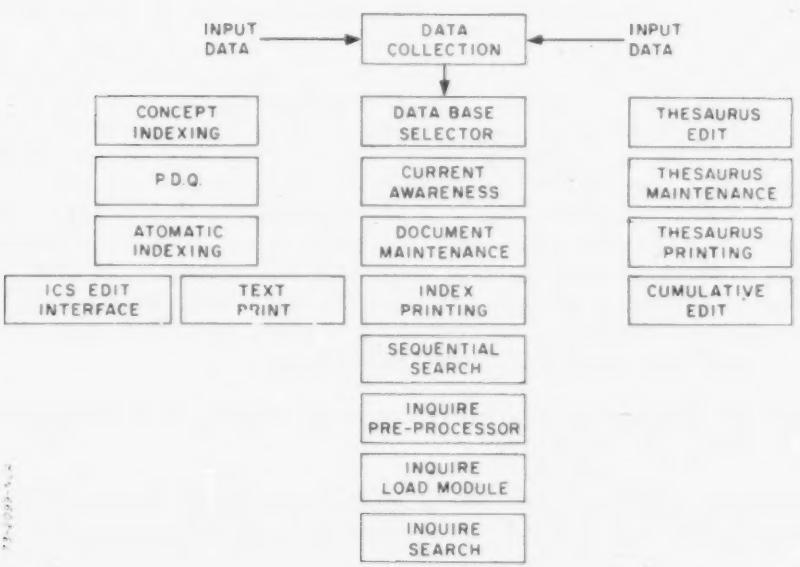


FIG. 1

We do not propose to discuss these modules in detail. However, we think these points are of interest: First, from a systems point of view, they can be regarded as standing alone. Suppose, for instance, some new development in either hardware or software suggests we might wish to make a change to the Document Maintenance Module. We can make it without touching any of the other modules.

Second, from a user-point of view, we can play all kinds of games. For instance, for any given information bank, we can elect to use concept indexing, P.D.Q. (our acronym for KWIC indexing) or automatic indexing, or we can combine types of indexing - we might, for instance, elect to put some deliberately selected concepts into an automatic indexing effort. We can build a thesaurus or by-pass that option. In fact, our options are almost limitless. However, once we choose a "path" through the modules for a particular task, the system reacts to it as if it was a single program.

The introduction of this degree of modularity into the system has had two major effects on our operations. First, it made it possible for us to develop our Microform Information Transfer Center, and second, it significantly lowered our operating costs.

In the M.I.T.C., which is shown schematically in Figure 2, COM-produced indexes are searched in a 16 m.m. cartridge reader-printer, and search results are retrieved in full-text from a microfiche file which is an integral part of the center. Because it is possible to perform Boolean searches of the indexes, and because search results can be delivered to the client as full-text, hard copy, with the help of the 16 m.m. fiche reader-printer, we occasionally and not entirely facetiously refer to it as our real-time, on-line, full text IR system. Its successful operation owes a great deal to our information bank philosophy and to the user being able to recognize which information bank is likely to contain the information he needs.

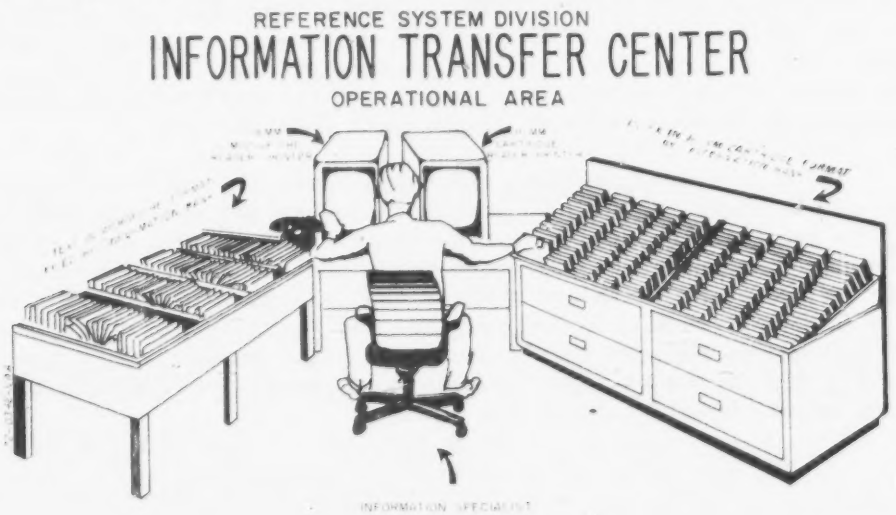


FIG. 2

Two comments before we leave the M.I.T.C. topic. One: - a COM-produced index costs 1/10th of its paper equivalent and is much easier to search. Second: - we can advance to real-time search on the computer as soon as demand for information and/or number of documents in any particular information bank indicate we can economically justify it. We merely begin to use the INQUIRE modules.

Now a word about costs. Operating with the system, it costs between 20 cents and $1.00 per document to update an information bank. The bigger the bank, i.e., the more documents it controls, the lower the cost per document. (How long this will continue to be true, we are not prepared to estimate.)

However, this is not too meaningful a figure, even if I tell you our computer charges are based on a cost of ± $800 per hour's CPU time, so let me put it this way. Operating with the IMM, we very quickly found that it cost us appreciably more for key-punching to get information into the computer than it did to manipulate that information once we'd got it in. It was this horrible fact of life that generated our interest in micro-computers - devices that offered the possibility of removing key-punch right out of our operation. Let me hasten to add that we quickly discovered they had much more to offer.

We make a subtle difference between micro-computers and mini-computers. A micro-computer can be called and regarded as an "intelligent terminal". It "stands alone" until the moment you want it to communicate with another and generally larger computer. The one we use has a built-in 12K memory (which is more memory than the second generation 1401's had).

Generally, it uses two cassettes, one to operate the programs we have developed for it and one to hold the information we are feeding into it. At any point of time, we can transmit the information on the second cassette, by telephone line, to our main computer some twenty-five miles away from our operating centre. When the main computer, a 370/165, has done whatever it is asked to do - it can transmit the results back to us and put the micro-computer back into business. We will return to that thesis in a moment but first of all we'd like to present some figures which we think might interest you.

At this moment, we have thirteen active Information Banks, and they control just over 27,000 documents. A recent analysis suggested that to get this control, we had to generate on average and let me emphasize "on average" 13.62 punch cards per document. The mini-computer we are using, with its printer and telecommunication link rents for ± $350 per month. If key punching and verification costs 10 cents per card and we believe that's about the going rate, then if you are entering 260 documents into the system per month, you pay for this rental merely by eliminating key punch.

The advantages of using a micro-computer, however, go further than this. Our indexers no longer have to fill in a computer-input-document. Rather, they examine a document, decide what key words they are going to use to index it, make a few notes and then feed the necessary information into the micro-computer.

We ran a time-test on this process recently. It suggested that on average "note-preparation time" took 4.19 minutes per document and "entry-time" took 5.50 minutes per document. If you figure that in New York an indexer is paid, including overheads and fringe benefits, $8.25 per hour or $1386 per month, then this test suggests that it costs about $1.30 to index a document and feed it into the micro-computer, ready for transmission to the main computer.

Let me sound a note of caution. These figures are based on N.Y. rates, and the type of documents we are working with. They are, we repeat averages. For comparative purposes, the same test indicated that using our old method of filling in an indexing input form took 7.85 minutes per document and of course, the information on the input form had to be key-punched before we could get it into the computer.

Now somewhat quickly we'd like to tell you a little more about how we operate with our micro-computer. It's not much bigger than an IBM Selectric typewriter and its keyboard is almost exactly the same. The main difference is that it has a T.V. screen attached, and it looks like this. (Fig. 3)

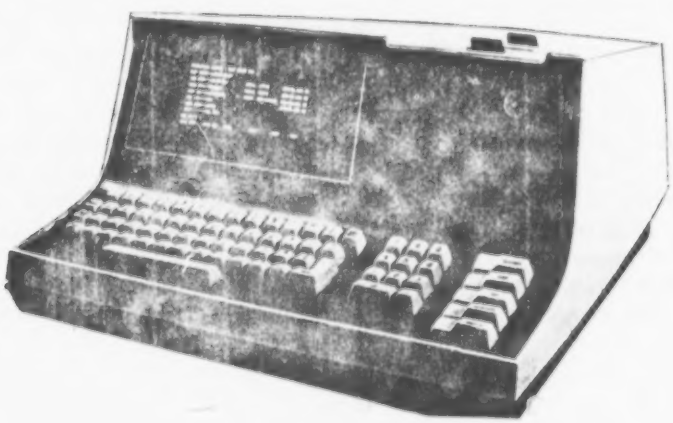


FIG. 3

Most people's first reaction to it is that the screen is very small. However, it can hold 960 characters - the same number as the very much larger IBM model 2260 CRT device.

The excellent character resolution and screen formatting capabilities built into soon remove this early impression.

We've set it up so that the user, having identified himself and been given certain pre-cautionary instructions, is shown what we call the "Menu" and it looks like this. (Fig. 4)

INDEX MANAGEMENT MODULE

1. CONCEPT INDEXING	6. RETRIEVE
2. DOCUMENT REVISIONS	7. PILOT
3. THES REVISIONS	8. D B UPDATE
4. P D Q	9. FILE SCAN
5. TEXT PROCESSING	10. TEXT EDIT

11. WRAP-UP

TO EXIT FROM ANY PROGRAM, PRESS & HOLD KEYBOARD
THEN PRESS ENTER KEY.

ENTER THE NUMBER CORRESPONDING TO THE
DESIRED FUNCTION .....ØØØ.

FIG. 4

In effect, the terminal is asking him what he wants to do. Now we would like to "for instance" it for you. Suppose he wishes to enter some Concept Indexing: he Keys-in 1 (at this point a cursor is blinking in the last position on the screen. .....ØØØ in Fig. 4). The terminal then searches for the appropriate program and in fact, shows him the equivalent of the Concept Indexing Input Form - line by line. When he has reached the 10th line, his screen looks like this. (Fig. 5)

CONCEPT INDEXING INPUT FORM

```
Ø1  D B
Ø2  DOCUMENT NO. ØØØØØØ
Ø3  D T C: ØØ
Ø4  LOGICAL SET OPTIONS:  (1Ø MAXIMUM).....
Ø5  DATE ØØ/ØØ/ØØ
Ø6  REFERENCE NO:
Ø7  TITLES 1 (T11):
Ø8           (T12):
Ø9  TITLE  2 (T21):
1Ø           (T22):

    ARE THERE ANY CHANGES:  ( Y OR N )
```

FIG. 5

When he has indicated that he does not wish to make any further changes, the screen format changes and he is shown how he should enter his keywords, get them into his thesaurus and "weight" them. Note the data is edited as it is entered and provision is made for making revisions to input data.

Obviously, time does not allow us to discuss our other options. We incline to the belief that you'ld like them. So let's leave this aspect of our micro-computer and just indicate one other thing about it which excites us.

We have found, to our great delight that it is a most efficient means of access to remote "public utility" information banks. For instance, we can access Lockheed's information bank in California and query the ERIC data base, the NTIS data base, and/or the Pandex data base at a price we can afford, i.e., $25 to $35 per terminal connect hour. Because we provide our own terminal and communications (through our tie-line system) this is the only charge we incur. Remember too, that the elimination of key-punch has paid for the micro-computer.

A HYBRID INTERACTIVE SEARCH SYSTEM
(UN SYSTEME HYBRIDE INTERDEPENDANT DE LA RECHERCHE DOCUMENTAIRE)

John D. Currie
Digital Methods Limited
1716 Woodward Drive
Ottawa, Ontario K2C OP8

ABSTRACT

Independent storage and retrieval systems based on either micrographic or computer techniques have been extensively used. A different approach is to assemble a flexible hybrid system in which bulk storage is in microform, while dictionaries used in searching are stored in a computerized data bank. The economics, response time and search methods of the Defence Scientific Information Service (DSIS) retrospective search system are examined. (Les systèmes de stockage indépendants et de recherche documentaire basés sur la micrographie ou sur les techniques de l'ordinateur ont été très employés. Un autre point de vue est de rassembler un système hybride flexible dont le stockage consiste en microcopies tandis que les questions formulées dans la recherche documentaire sont rassemblées dans une banque de données. L'économie, la vitesse de réponse et les méthodes de la recherche documentaire du Service d'Information Scientifique pour la Défense (SISD) sont ici examinées.)

INTRODUCTION

Independent storage and retrieval systems based on either micrographic or computer techniques have been successfully used for some time. This paper examines the merits of a flexible hybrid system that is being developed in the Defence Scientific Information Service (DSIS) of the Defence Research Board (DRB).

Since 1946 DSIS has been responsible for collecting and cataloguing Canadian and foreign scientific and technical material of defence origin and disseminating this material to those who participate in Canadian defence and development. Much of this material is of a classified nature.

Since 1968 DSIS has recorded document accessioning data in machine readable format. These data are then used to generate various types of announcement, including KWOC, Digests and SDI and are also used to produce conventional type catalogue cards.

BACKGROUND

Retrospective searches until recently have been conducted using manual methods primarily centered around catalogue cards. With the availability of machine readable data retrospective searching can be automated.

The total DSIS document collection includes approximately 350,000 titles of which only 40,000 are in machine readable form, i.e. accessions since 1968.

The costs of entering data in the same readable form, i.e. full citation and abstract have precluded adding older documents (pre 1969) to the computer files. However, to improve searching efficiency, to reduce storage requirements and to allow the older document collection to be more accessible to DSIS customers it was decided to microfilm all citations (catalogue cards) relating to the older documents. This has allowed the older material to be subsequently included in the computer files.

Microfilming has been completed for some files (subject) and involved re-ordering of the cards to suit current DSIS indexing practice. A further film ordered by accession number has also been filmed.

Data from the subject films has allowed DSIS to provide machine retrospective searching for its total collection. For the older documents (pre 1969) the accession numbers associated with a particular subject heading have been manually read from the subject ordered film and input to a computer file. Thus the total collection can now be searched by subject and more recent documents by other surrogates. A computer search can be used to output actual document citations for recent accessions and a microfilm cross reference for older material.

The original plan for the machine readable data was to store the complete bibliographic data on files with controlled access to users via secure telephone lines. Because of the current unavailability of a secure communications network and of Department of National Defence computing facilities which could provide the type of interactive query service we require, an interim hybrid system has been developed. An unclassified data base containing the data is stored in a public computer in Toronto. Each user terminal of the system has a microfilm reader associated with it, together with a set of microfilm cartridges containing catalogue card images of all accessions in the DSIS document collection. The images are in accession number order. The results of a computer search indicate the frame numbers at which the full bibliographic data is stored on the film. Viewing the abstract allows the user to determine whether he wishes to see the document.

SYSTEM ORGANIZATION

Use of this system provides almost the same search capability previously available only at DSIS, Ottawa. The use of the microfilm allows citations associated with classified material to be made available to the end user even though the index system employs a public utility and normal public communications facilities (voice-grade telephone).

SYSTEM ORGANIZATION

The DSIS retrospective search system consists of three sub-systems:

(a) The on-line terminal connected via a telephone line to the computer based index material.

(b) The microfilm sub-system consisting of a viewer and microfilm cartridges which are referenced by the computer data.

(c) The actual documents which will in many instances be available in microfiche form in a location convenient to the on-line terminal.

Terminal and Computer Sub-System

This sub-system consists of three major components: the terminal; the computer system and associated programs; and the data base.

Terminals. Currently the Bell VUCOM I and the IBM 2741 terminals are used in the DSIS system. The VUCOM I terminal consists of a keyboard attached to a Cathode Ray tube (CRT) device which displays the keyboard entries and the computer responses. An optional printer is available which can be used to print out (hard copy) of selected items which are also displayed on the CRT. The IBM 2741 terminal uses a standard IBM Selectric typewriter. Keyboard entries and computer responses are recorded on normal fan-fold continuous stationery. The VUCOM system operates at a higher speed (30 characters/sec versus 15 for the 2741) and is noiseless. Its major disadvantage is limited screen size. This can be overcome by using the printer, but at increased cost.

Computer Systems and Software. The DSIS system uses the computing system available from IP Sharp Ltd. Their computer is located in Toronto and can be accessed via the public telephone systems from specific locations across Canada and the USA with IP Sharp absorbing the long distance telephone charges from these cities. (Vancouver, Calgary, Toronto, Montreal, Quebec, Halifax, Washington D.C., etc.)

SYSTEM ORGANIZATION

The IP Sharp system employs an IBM computer, Model 370/145 which includes extensive on-line storage. The computer system operates in a timesharing mode and provides a service known as 'APL-PLUS'. This service allows the high level programming language APL to be used simultaneously by many users. Each user is serviced so rapidly by the system that under most operating conditions the system appears to be dedicated to the individual user.

The DSIS retrospective search system uses the APL-PLUS system at the same time other users are carrying out their own computations. In addition, many users may be using the retrospective search system at the same time, each user having his own copy of the programs associated with the system, but all sharing the same data base. (Index material). When a user signs onto a system he is given a copy of the programs in an area of computer storage known as workspace. The workspace and the copy of the programs are then dedicated to his use all time he is signed onto the system.

The search system can be used by library clerical staff or scientists with a minimum of training since the query language involves simple English questions and answers. The system is also being developed in a bilingual version. Various levels of user expertise in the system are accommodated by providing a conversational (prompted) mode and a command (unprompted) mode.

The retrospective search system has been designed and developed by Digital Methods Limited for DSIS.

Data Base. The data base used in the retrospective search system consists of a series of files which are used by the system to select documents which meet the requirements of the user's search questions. The files are organized in a similar manner to the conventional library catalogues and allow searches to be conducted alphabetically by such terms as personal author, corporate author, descriptors and keywords, and COSATI code.

Data relating to documents included in the system are organized in the form of indices. These indices are searched separately, e.g. subject or author, but the results of an author search can be logically combined with the results of a subject search to find documents written by a particular author on a particular subject. The logical functions AND, NOT and OR are available.

The results of a search on the data base are displayed to the user in stages, first he is told how many documents have been found which meet his requirements, e.g. a certain subject or author. The actual data base in the system relating to the documents can then be requested.

Thus by refining the search the user can limit the final results to a small set of documents which meet his particular requirements.

The actual data about a particular document which is displayed to the user is limited to the title and various cross references. This approach has been taken to allow all DSIS material including classified documents to be available from the system. The cross reference material relates to:

(a) The original DSIS accession number.

(b) A microfilm number which allows the full catalogue and indexing material to be selected from the microfilm.

(c) A microfiche number which will eventually allow the document microfiche to be selected from the library file.

Microfilm Sub-System

Microfilm Cartridges. Cartridges provided by DSIS for the retrospective search system have been produced by filming existing catalogue cards for the years 1946 to 1968 inclusive and for the years 1969 to date by computer output microfilm (COM). In both the manually produced film and the 'COM' there is a small square 'blip' encoded opposite each frame of film for ease of retrieval. The cross references provided by the computer system identifies the cartridge number, the frame number, and the accession number of the document.

Microfilm Reader. In order to take full advantage of the computer indexing and the film format in the microfilm cartridge it is essential to have a good reader-printer. DSIS use a page search reader-printer which functions by counting the blips encoded on the film. If the proper cartridge is inserted in the frame and the correct frame number is keyed in the reader automatically stops at the selected frame so that the desired accession number as specified by the computer cross reference data is available for viewing. Retrieval of a particular document catalogue card image including loading the cartridge is of the order of seconds.

Microfiche Sub-System

The ideal microfiche support for the DSIS retrospective search system is a microfiche library of all documents accessioned by DSIS at locations of customers that have terminals. As an interim goal master microfiche of all documents accessioned since 1969 are being made available at DSIS. Duplicate copies for customer's requests will be provided. At a later stage complete or partial sets of these fiche may be provided at the customers terminals.

SYSTEMS COSTS

This time-sharing system provides the user with a very powerful system at low cost since the user's cost in addition to terminal rental and data base storage reflect only the actual individual use made of the computer processing time. The costs associated with the individual user are made up of two components:

(a) The time connected to the system.

(b) The cost associated with the actual computing time, i.e. the total of all the small increments of processing time during which the user actually has the computer working for him.

Service Charges. These are:

connect time	- 20 cents per minute
computer time (CPU)	- 30 cents per second
terminal rental (VUCOM + printer + acoustic coupler)	- $200 per month
data base storage	- $20,000 per year

Assuming three searches per day at each of 10 terminals with the average search being fifteen minutes connect time and fifteen seconds computer time. The estimate cost per search is:

connect and CPU time	- $7.50
terminal rental and data base storage	- $5.50
Total cost per search	$13.00

The estimates for search time are based upon our limited experience and will no doubt be revised.

The total cost of producing the microfilm used in the retrospective search system for DSIS and all Establishment libraries was approximately $7000. Microfiche readers vary in price from $1,500 to $5,500. We use the more expensive page search reader-printer, however this equipment was already in use in a stand alone microfilm catalogue system made available to our major customers. This system allowed manual search of microfilm after identifying relevant subjects in a hard copy index which provides reel and frame number references.

CONCLUSION

To include the bibliographic data in the computer file rather than on microfilm would more than triple the actual storage costs, i.e. $65,000 rather than $20,000 and in addition would substantially increase the computer costs for each search. To this must also be added to cost of encrypting equipment at both the computer and the remote user. (Obviously a secure classified computer installation would also be required.)

CONCLUSION

What we have achieved is not our ultimate goal, but it has permitted us to provide an improved service to our customers in a relatively short time frame at reasonable cost and at the same time overcome problems of security that could have delayed our system for an unacceptable length of time. We are, in effect, using the media to their best advantage: the computer for the manipulation of information and the microfilm for storage of information.

REFERENCES

AMEY, G.X., 1971, "Use of Simulation in Design of an Online Information Retrieval System." IEEC Conference Digest, October 1971, p. 60-61.

_____, "Evolutionary Transition to Automation of Information Management." The Fifth Annual Colloguium on Information Retrieval: Ramey J.A., Information Interscience Incorporated, p. 63-70.

_____, 1969, "A National Information Utility for Communication of Scientific Information." IEEC Pre-Conference Digest, October 1969, p. 108-109.

_____, 1968, "SOCRATES - An Information System for the Canadian Defence Community." Proceedings of Session on Information Retrieval at the 1968 Defence Research Board Symposium.

CURRIE, J.D., 1972, "Microforms in DSIS: An Information Center's Experience with Microforms." DSIS Technical Note No. 9, July 1972.

OPTICAL CHARACTER RECOGNITION IN INFORMATION PROCESSING
(RECONNAISSANCE OPTIQUE DES CARACTERES IMPRIMEES INFORMATIQUE RESUMES)

Ching Y. Suen
Sir George Williams University
Montreal 107, Quebec

ABSTRACT

The use of optical character recognition (OCR) speeds up the flow of processing printed information. This paper considers the application of OCR to the direct scanning of documents. Its characteristics and capabilities are examined. Factors affecting the recognition rate and methods of improving the performance are studied. The recognition of constrained handprints and the human factors involved are discussed. (L'utilisation de la méthode de reconnaissance optique des caractères(ROC) acceleré la procédure des données imprimées. Cette étude montre les applications de la méthode ROC dans le déchiffrage direct des documents. Les caracteristiques capacitées de cette méthode ainsi que les facteurs affectant la vitesse de reconnaissance et les méthodes d'amélioration de la performance y sont étudiés. Aussi la reconnaissance d'empreintes manuelles contraintes et les facteurs humain concernés y sont éxaminés également.)

INTRODUCTION

Optical character recognition (OCR) is the conversion of printed or written information into computer readable codes for subsequent processing. Its development stemmed mainly from the rapidly increasing volume of data to be processed by the computer.

The first OCR machine appeared in 1954. Despite their large number of potential applications, optical readers did not attract widespread interest in the nineteen sixties. High costs and limited capabilities were the two chief reasons. Most of these machines cost between $150,000 and $1,500,000 and could read only one or two specific character fonts. Acceptance of these earlier models of optical character readers was further hampered by the availability of keypunch machines and the ever-increasing number of keypunch operators.

Fortunately, technological advances in recognition techniques and digital electronic equipment in the past few years have cut down the cost of OCR components considerably and at the same time made possible an increase in OCR capacity. Most modern OCR machines can read at a high speed several common typewritten character fonts as well as some handprinted

symbols and numerals. As a result, the population of OCR machines has expanded rapidly. It has become one of the most promising techniques for overcoming the bottleneck of data input to the computer. Its applications include data capture of cash register tapes, sales slips, credit-card sales tickets, utility stubs and meter readings, bills and invoices, medicare claims, mail and a great variety of data records and commercial forms.

OPERATION OF THE OCR SYSTEM

A block diagram of the OCR system is shown in Fig. 1. Operation of the entire system is controlled by a console which consists of a teletype or a video display unit. Documents containing the required information are first fed into the document transport unit which in turn conveys each of them through the scanner. Flying spots and photocell arrays are the two most common types of optical scanners. The function of the recognition unit, that is, the digitization and recognition of characters, is based on the scanning of the contrasts in patterns of the printed characters against the paper background. The recognition process consists of the extraction of distinctive features of the character and matching them against tables containing the characteristics of a trained set of characters. Processed characters are classified into recognizable and unrecognizable groups; documents containing the latter will be directed into a reject stacker for subsequent amendment. In addition, numerous OCR systems are now equipped with editing facilities. Documents containing unrecognizable characters can be retrieved immediately and the operator can enter from a keyboard the correct data into the recognition unit. In some cases, a video display is used to show the processed characters on a screen. Unrecognized characters can thus be corrected on-line. Once a document has been read, the information it contained will be transferred into a storage device for further processing. Magnetic tapes, discs and paper printouts are the three most commonly used storage media.

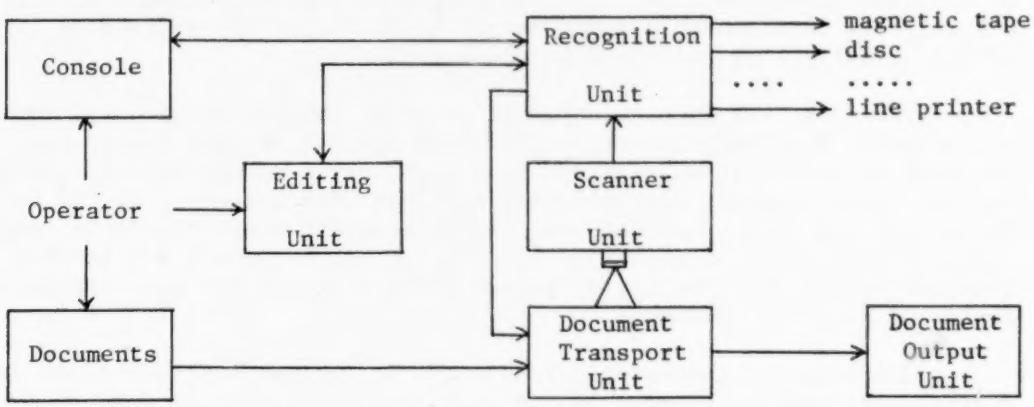


Fig. 1 Block diagram of the OCR system.

OPTICAL CHARACTER RECOGNITION

OCR CAPABILITIES AND CONSTRAINTS

Incorporating the recent advances in electronic technology and pattern recognition techniques, modern OCR machines can read a number of common typewritten fonts. To aid the manufacturers and the users, standdard OCR fonts have been developed; among these are the OCR-A font developed by the American National Standards Institute (1966) and the OCR-B font designed by the European Computer Manufacturers Association (1971). These standard fonts which have gained widespread acceptance, aim at achieving optimum machine readability while at the same time taking conventional character appearance into consideration. Style A emphasizes machine reading performance while style B emphasizes conventional appearance. Both styles include upper and lower case letters of the alphabet, numerals and special symbols. In addition to these standards, a number of fonts have been developed by individual manufacturers. Some machines in the market can read the OCR-A font only; others can read not only a variety of common machine printed fonts, but also handprinted numerals and a few special symbols. OCR machines which have only one reading head can process only one line of information at a time; those equipped with more than one reading head can process a corresponding number of lines at the same time. Error rates as low as 0.001 - 0.003% have been quoted by manufacturers. The average size of documents handled varies from 3.0" X 3.5" to 4" X 8"; some machines can even accommodate full size documents of 8.5" X 11.0". The rate at which documents can be processed depends on the amount of information carried, the design of the documents, the number of lines of characters and the spacing between them. The processing speed can vary from 20 up to 3,600 characters/sec. (Andersson, 1971).

Apart from its capacity of processing data, the optical reader can also be applied to read books and journals, digest papers, extract index terms and abstract articles. Here one faces more constraints than reading typewritten materials. First, the optical reader must be equipped with a dictionary having a large vocabulary of words. Second, it must have a multipurpose transport unit to handle the source materials properly. Third, it should be capable of reading virtually an unlimited number of fonts and characters of many different sizes. Microfilming is one way of alleviating some of the above problems and machines which can read microfilm materials are now available.

In order to facilitate recognition, documents must be prepared very carefully, and a number of constraints must be observed. The most important factor affecting the recognition rate is the formation of characters. To enhance the recognition rate, characters must be clean and sharp, have a uniform density and be free from voids and black-in spots. They must also have a low reflective index with sufficient contrast against the paper background. Similarly, the paper itself must be clean, smooth, opague and highly reflective. Also, paper stiffness and thickness

must be designed to maintain durability and to facilitate its conveyance by the transport unit. High-quality carbon ribbons are generally recommended. Typewriter keys must be clean. Apart from the above, the stroke width of characters, character separation and rotation, line and field separation, must all conform to specifications. Some manufacturers publish instruction booklets to help users in the design of documents to optimize OCR efficiency.

Type font variations do not generally present any difficulty to human character recognition. This is because understanding is facilitated by reading the characters in their context. In the case of OCR, machine recognition of characters depends solely on the analysis of their distinctive features. Although some machines have a learning capability to adapt to reading new fonts, the difficulty of recognizing the more than a hundred existing type font styles is self-evident. As a result, many machines are programmed to read only one or two type fonts. This places severe restrictions on many users since documents generally come from a variety of sources. A relatively high rejection rate can also be expected from documents which have undergone a large amount of handling due to the accumulation of dirt smudges and creases.

ECONOMIC CONSIDERATIONS

In determining the economic feasibility of employing OCR as a means of processing data, several factors must be taken into consideration. These are the volume of data, throughput speed, character fonts, document size, form design; and above all, its cost as compared with other methods of inputing the data.

Keypunching is probably the greatest rival of OCR in information processing and several reports have been published describing their relative merits, see for example, Andersson (1971), Tierney (1972) and Sheinberg (1968). These analyses indicate that OCR is economically feasible when the amount of data to be processed exceeds half a million cards. The cost for processing the above amount of data by the keypunch method is equivalent to the employment cost of about 9 to 12 full-time keypunch operators plus the rent for keypunch machines and their acessories. With increasing labour costs, rising volume of data and decreasing costs of digital equipment, the break-even point will be considerably lowered in the future. Since OCR minimizes manual intervention of the recorded data, it is therefore less susceptible to transcription errors.

The high efficiency of OCR has the advantages of quick throughput, short billing cycle and quick information and sale transfer. As a result, OCR has attracted a large number of users such as utility companies, banks, post offices, government agencies and other large enterprises. These organizations which handle large volumes of data have reported immense savings through the use of OCR.

OPTICAL CHARACTER RECOGNITION

RECOGNITION OF HANDPRINTED CHARACTERS

The ability of reading handprints is a desirable feature in OCR since it can eliminate the expensive and error-prone process of data transcription. This capacity is especially useful in the many circumstances where typed documentation is either impossible or impracticable, e.g. outdoor data collecting and meter reading. It is particularly useful in processing handwritten programs.

Although many OCR machines can read multifont and intermixed font characters, they can read only a few handprinted characters. Since each writer has his own style of writing, handprints are less tractable than multifont characters and recognition of the former must be based upon general character configurations rather than simple matching processes. Current recognition techniques have been described by Harmon (1972).

One way of facilitating handprint recognition is to develop a standard such that the person entering the data can write in a style as close as possible to the given standard. One such standard is now in preparation at the American National Standards Institute (1972). Some manufacturers have also produced their own set of stylized characters. This includes all the numerals and special characters C, S, T, X, Z, +, - and /.

To facilitate machine recognition and to aid the writer in printing characters suitable in shape and size, pre-printed boxes are used as guidelines. The user is instructed to write the characters neatly within these boxes with a sharp black-lead pencil (e.g. HB pencils). Pre-printed boxes have a pitch of 4 to 5 characters per inch, their sizes vary from 0.16" X 0.20" to 0.24" X 0.32". There is also a limit on the stroke width, normally between 0.01" and 0.04". The slopes of both vertical and horizontal lines should also conform to specifications. The handprints should resemble closely the pre-printed samples, otherwise a high rejection rate can be expected.

Because of the rigidity of these rules and the difficulty of changing one's established style of writing, the writer usually has to undergo some kind of training process before he can print characters acceptable to the machine.

Through motivation and patience in training, successful recognition of appropriate printing styles of numeric handprints can be achieved. This has been reported by a number of organizations in diverse applications such as payroll, accounting, driver's license registration and meter reading.

OPTICAL CHARACTER RECOGNITION

HUMAN FACTORS INVOLVED

OCR machines are very sensitive to print degradation and document mutilation. In order to cut down the error and rejection rates and to improve the overall performance, co-operation of the users is necessary. This is particularly so in handprint recognition.

The results of a handprint experiment carried out at Cognitronics Corporation (1972) revealed that character deformity was the chief reason for rejects. Other factors include interference between over-sized characters, improper erasures, stray marks, etc. It was also shown that a large proportion of rejects was generated from the handprints of a small number of tested writers. In an experiment recently conducted by Suen (1973), approximately 2,000 handprinted characters were collected and analyzed by an OCR reader. The results indicated that characters written by subjects with a light stroke yielded a recognition rate 4.7% lower than those written with a heavy stroke. Erased characters yielded a recognition rate about 10% lower than those without erasure. The above results suggest that more research should be done to investigate human behaviour in handwriting. In order to avoid ambiguity in man-to-machine communication, harmony between machine specifications and human handwriting practices should be reached.

OCR PROSPECTS

Many OCR machines have demonstrated capabilities markedly superior to earlier models. Improvement in recognition techniques and logic designs, accompanied by the reduction in cost of components, have stimulated the growth in the use of OCR readers. Although numerous human factors remain unsolved in OCR recognition of handprints, satisfactory training programs have been developed by both manufacturers and users, attesting to the progress in handprint recognition.

With the continuous effort aimed at character and document standardization, both nationally and internationally, it is expected that a powerful impetus will be given to the introduction of more low-cost, special — purpose OCR machines. There is little doubt that OCR will become one of the most prominent technologies in information processing in the next twenty years.

REFERENCES

American National Standards Institute, Inc., "USA Standard Character Set for Optical Character Recognition," Standard X3.17, (1966).

_____, "Proposed American National Standard Character Set for Handprinting," Document no. X3A1/72-60, (1972).

OPTICAL CHARACTER RECOGNITION

Andersson, P. L., 1971, "OCR Enters the Practical Stage," Datamation, v. 17, no. 23, p. 22-27.

Cognitronics Corporation, "Handprinted Digit Reading under Conditions of Limited Control," Nov. (1972).

European Computer Manufacturers Association, "Standard ECMA-11 for the Alphanumeric Character Set OCR-B for Optical Recognition," Oct. (1971).

Harmon, L. D., 1972, "Automatic Recognition of Print and Script," Proceedings of the IEEE, v. 60, no. 10, p. 1165-1176.

Sheinberg, I., 1968, "Optical Character Recognition for Information Management," in: Kanal, L. N., ed., Pattern Recognition: Washington, D. C., p. 31-40.

Suen, C. Y., 1973, "Factors Affecting the Recognition of Handprinted Characters," in preparation.

Tierney, D. F., 1972, "Traditional OCR Problems Passé in Today's Market," Computerworld,(Nov.), p. 8.

A UDC WATER THESAURUS/CONCORDANCE: DEVELOPMENT AND USE. (LA CONCORDANCE DE CDU AVEC UN THESAURUS DE L'EAU: DEVELOPPEMENT ET UTILISATION)

Helen E. McCuaig
Marcel A. Mercier
Environment Canada
Ottawa, Ontario K1A 0E7

ABSTRACT

The efficient operation of an automated retrieval system requires that information in the system be effectively identified. Identification may be made through a controlled vocabulary, an uncontrolled vocabulary, or a classification or coding scheme. Each method has unique deficiencies which do not occur in others. With the advent of more sophisticated computer packages for information retrieval, it is possible to combine various identification methods and thus cancel individual disadvantages. In certain instances the system's vocabulary must be linked with the classification. A study to achieve such linking has been conducted at Environment Canada and has resulted in the concording of a water resources vocabulary with the Universal Decimal Classification scheme. Problems that arose, various tools that were used, basic methodology followed and results of the work are given. Plans for utilization of the concordance for indexing and information retrieval purposes within the WATDOC System are discussed and its potential use as an international switching language is outlined. (L'efficacité d'un système automatique de recherche documentaire requiert une identification claire de l'information. L'identification se fait par un vocabulaire contrôlé, un vocabulaire libre ou un système de classification ou de codage. Chacune de ces méthodes présente des déficiences uniques. Mais, avec l'avènement accru de programmes d'ordinateurs sophistiqués reliés à la recherche documentaire, on peut combiner diverses méthodes d'identification et éviter ainsi les désavantages particuliers de chaque méthode. Pour arriver à cette combinaison, il faut relier le vocabulaire utilisé au système de classification. Une étude afin de réaliser cette combinaison a été effectuée au ministère de l'environnement du Canada.

WATER THESAURUS/CONCORDANCE

Une correspondance a été établie entre un thésaurus de l'eau et le système de "classification décimale universelle". L'article suivant présente les problèmes rencontrés, les instruments de travail utilisés, la méthodologie de base suivie et les résultats obtenus. Les perspectives d'utilisation de la concordance "EAU/CDU" reliées à l'indexation et à la recherche documentaire à l'intérieur de WATDOC y sont discutées, ainsi que les possibilités d'utiliser la concordance comme langage de commutation.)

INTRODUCTION

During recent years information systems have increasingly dealt with matters of a multidisciplinary nature and increasingly have been directed towards a mixed clientele. WATDOC, Water Resources Document Reference System, Environment Canada, is an example of such a system, where references to all types of environmental information related to water resources are acquired from and used by the participants. The users are found at all levels of management and research in both the public and private sector.

For the efficient operation of such a system, it is frequently necessary to identify the information within the data base on a conceptual level and then within the general concept to separate documents through precise semantic coupling of keywords. Ohman (1971), Aitcheson (1970). WATDOC has attempted to provide this facility through the development and subsequent use of a concordance between a water resources thesaurus and a widely used classification scheme, the UDC.

With the tools that are available today, special editions of general classification schedules can be developed from the universal set and then the special edition may be combined with a structured vocabulary. In the case of an international classification it is possible to use such schedule modification for the development of an international switching language and to provide also an interdisciplinary switch between systems. Lloyd (1972).

It is interesting to note that the latest UNISIST recommendations indicate the need for a new broad system of ordering, that is a general classification, to identify information effectively and to allow its movement between various systems. However, despite such recommendations and despite extensive work on classification theory, Wilson (1972), people have not yet come to grips with the complex problems posed by classification systems. Some of these are the need for varying levels

of ordering, for linking relationships, and for a truly universal approach to classification. UDC, in spite of its deficiencies, is the only widely used scheme that attempts to provide any such facilities. Scribor (1971), Stueart (1971), Saveljeva (1971), Rigby (1970), Wright (1969).

CHOICE OF CLASSIFICATION AND INDEXING SYSTEMS

The Universal Decimal Classification was selected for the project of Environment Canada (WATDOC), on the basis of studies completed at the University of Alberta. These studies, even though preliminary, demonstrated that UDC could be regarded as a metalanguage and used effectively for automated retrieval purposes in water resources management. Mercier (1971). Through the computer manipulation of machine readable data bases, a significant correspondence between water resources terms and UDC notations was found.

To meet the requirements for controlled indexing within the WATDOC system it was recognized that a thesaurus was needed. The thesaurus would provide control of the vocabulary and would allow for consistency in selection of indexing terms; it would be a tool used by both indexer and searcher. Wellish (1972) The U.S. Water Resources Thesaurus was a natural choice.

COMBINATION OF SYSTEMS

The concordance between a Canadianized version of the U.S. Water Resources Thesaurus and a subset of the Universal Decimal Classification is to be used as an indexing, classifying and retrieval aid. The concordance and the subset of the UDC are being developed in parallel. Tapes used contained the U.S. Water Resources Thesaurus, sections of the UDC schedules, and the AIP/UDC magnetic tape. Mercier (1971, p.28).

Universal Decimal Classification Scheme

UDC is a system of classifying information by analysis of idea content so that related concepts are grouped and subordinated. It is a hierarchical, numerical classification based on the Dewey principle that all knowledge is a whole and can thus be designated by 0., with infinite capability of subdivision as a decimal fraction. It is very flexible in that it identifies certain recurring faceted features such as form, time and point of view. These, indicated by auxiliary numbers, can be appended to any concept. Other types of synthesis can also be indicated. British Standards Institution (1963), Mills (1964), Foskett (1969).

UDC is under the auspices of an international body, Fédération Internationale de Documentation (FID). Its tables -- though not all

parts -- are in 22 languages. Abridged editions have been published and are useful for small collections. Several Special Editions exist for the fields of Metallurgy, Nuclear Engineering, Education,and Polar Studies, where all relevant UDC numbers are brought together, those for the core subject in full detail and those on marginal subjects in abridged form. It is anticipated that a Special Edition for Water Management, in machine readable form, will grow out of this project. This water edition will exist as a concordance to the water resources thesaurus. Lloyd (1972). The Canadian National Committee (CNC/FID) has recently been enlarged and although the UDC has not been widely known or used in Canada, several interesting projects are underway.

U.S. Water Resources Thesaurus

The thesaurus used in this project as a basis for the core retrieval language was developed by Water Resources Scientific Information Center, U.S. Department of the Interior. It consists of 4,288 main terms and cross references interfiled with USE references. The thesaurus was "patriated" somewhat by deleting from it certain specific American terms, mostly geographic, and appending to it additional terms used extensively in Canada.

PROBLEMS IN DEVELOPMENT

In carrying out the project, the major problems were the acquisition of up to date UDC schedules and the lack of staff expertise in UDC.

Acquisition of Schedules

During the first phase of the research,difficulty was experienced in obtaining up to date schedules. This was largely overcome through the use of the magnetic tape containing the complete merged set of UDC schedules prepared by the American Institute of Physics Universal Decimal Classification Project, 1965 to 1967. Freeman (1968). For the present phase the full Special Schedules and the Abridged English Edition were used in printed form. However, there were many difficulties in acquiring all schedules* and in order to get on with the task schedules were borrowed, or photocopies obtained from the National Library, National Science Library (Building Research) and International Development Research Centre.

---

* In Canada all orders for English or French editions should be made through the Canadian Standards Association.

WATER THESAURUS/CONCORDANCE

Special editions were very helpful in this project; a list of those referred to were:

(a) Chiapetti, F. et al, 1971, An International Language for Information Retrieval: The UDC-Thesaurus in the Water Field C.N.R. Institute di Recerca Sulle Acque, Rapporto Interno 18, Roma.

(b) Deutsche Dokumentations-Zentrale Wasser (DZW), 1971, Ordungssystematik zur Dokumentation Wasser: Begriffsverzeichnis: Thesaurus, Erich Schmidt, Verlag, Berlin.

(c) Fédération Internationale de Documentation, 1955, ABC Abridged Building Classification, 2nd ed., Beuweentrum, Rotterdam (FID 261)

(d) Fédération Internationale de Documentation, 1963, Universal Decimal Classification for use in Polar Libraries, 2nd ed., revised, S.P.R.I. Occasional Paper No. 2 (FID 348)

(e) Food and Agriculture Organization, 1950, Handbook for World Fisheries Abstracts, Washington, D.C.

(f) International Committee for Atmospheric Sciences, 1966, ICAS Vocabulary (Preliminary Edition), Washington, D.C.

(g) National Research Council of Canada, 1965, Index of Report Files in the Building Research Library, compiled by Mara E. Karnupe, Technical Paper No. 197, Division of Building Research, Ottawa.

(h) U.S. Department of Housing and Urban Development, 1971, Urban Vocabulary, Washington, D.C.

(i) New schedules for Classes 551.48/.49 (Surface and Groundwater Hydrology within the Geomorphology Class) and Class 556 (General Hydrology) which have been acquired through correspondence with researchers in the field.

Lack of Staff Expertise in UDC

The individuals who participated in this phase of the project were a librarian and a professional assistant. No major problems arose although they were not experienced in the use of UDC. Programmed texts by Wellisch (1970) and Perreault (1969) were used to acquire familiarity with the principles and techniques of UDC.

WATER THESAURUS/CONCORDANCE

DEVELOPMENT OF THE CONCORDANCE

Phase 1

Three copies of the U.S. Water Resources Thesaurus were used, two as working copies and one as the master. The first step was to transfer notations derived in the Edmonton studies, Mercier (1971), to the second edition of the U.S. Thesaurus. Following this two major tasks were necessary. These were the further verification of the machine concorded notations derived in the Edmonton study and the manual assignment of notations to those descriptors not previously concorded. The basic steps are depicted in Figure 1; they are described as follows:

Step 1: The Abridged edition was checked to verify existing descriptor notations and as an initial step in assigning new notations for those descriptors previously left blank.

Step 2: If Step 1 was unsuccessful, the other tools on hand were utilized.

Step 3: If a match could not be made, an attempt was made to assign a notation to the term by synthesizing two UDC classes, or adding auxiliaries, otherwise no number was assigned.

Step 4: All notations were neatly written into the Master Copy for encoding. When completed, the various sections of the Master Copy were passed on to the keypunchers for keying. In this manner all steps were completed in parallel.

Phase 2

All terms and existing notations were punched onto cards so that various lists could be compiled and printed. The basic steps were as follows:

Step 1: The terms and their notations were punched in fixed fields.

Col. 2-35	Descriptor
Col. 47-68	UDC notation. If no notation occurred the field was left blank.

WATER THESAURUS/CONCORDANCE

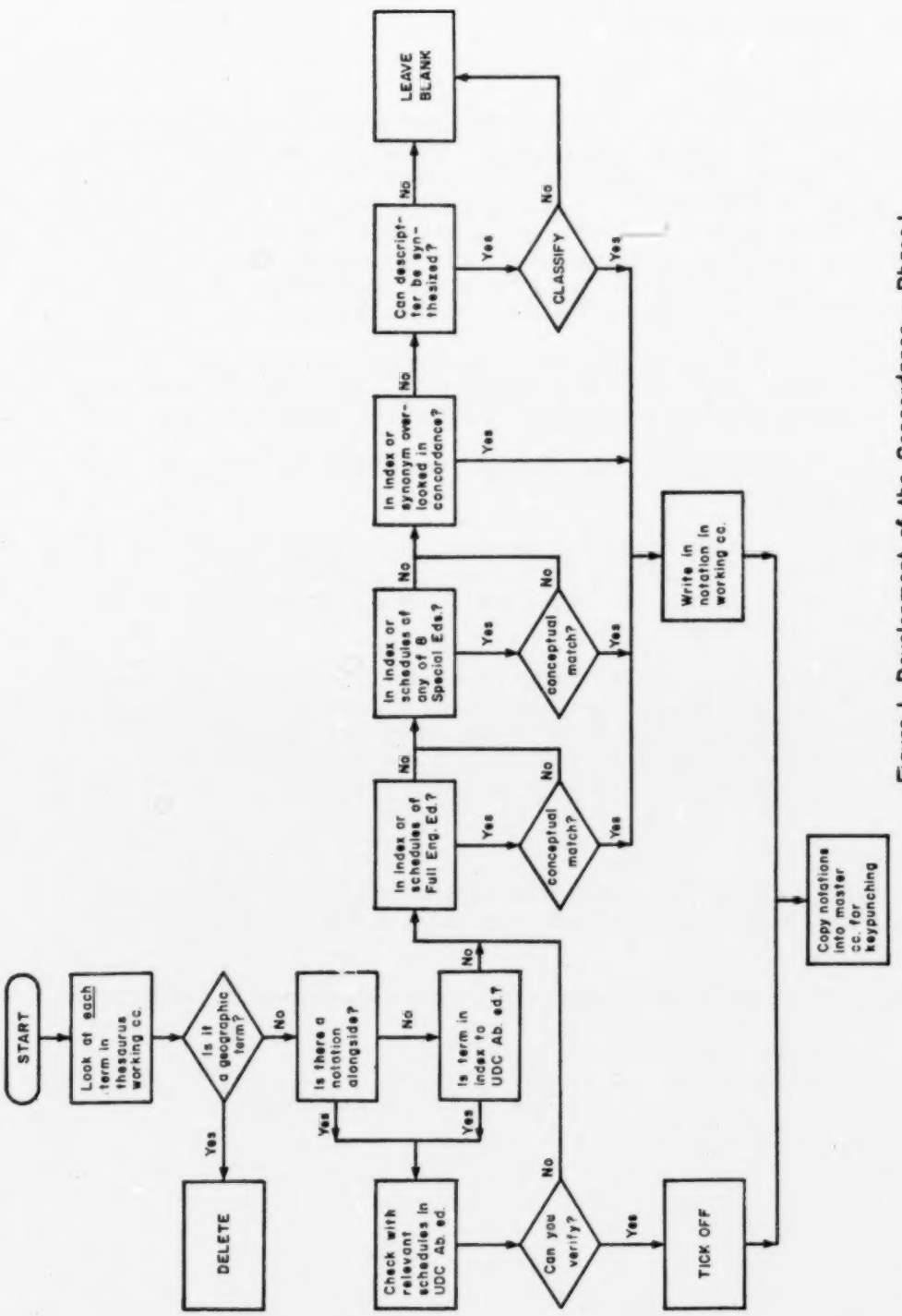


Figure I. Development of the Concordance - Phase I

Step 2: The cards were copied onto tape for ease of computer manipulation. Two listings were then produced:

1) An alphabetic listing
2) A numerical listing by UDC term

Both listings consisted of the descriptor term and the UDC notation. A sample of the alphabetic listing is given in Figure 2.

Step 3: For clarity, headings were added to the numerical listing. This was done by punching two cards for each heading.

Card 1,	Col. 16	*
	Col. 47-48	UDC notation
Card 2,	Col. 16-45	Heading
	Col. 47-48	UDC notation

These cards were added to the previous file. A sort was made and a numerical listing, containing headings, was printed. The headings were based on those occurring in the abridged and full editions of the UDC. They were selected on the basis of frequency of groups of notations occurring in the listing. A new heading was also inserted if a notation seemed remotely connected to the nearest heading. A sample of the resulting list is given in Figure 3.

Phase 3

To date, phase three of the project has not been completed. The object is to produce a printed subset of UDC schedules in machine readable form. Three basic steps have been defined, which are presently in parallel operation. These steps are as follows:

Step 1: Copies of the alphabetic and numerical listings and the UDC schedules have been distributed to professionals, within the Department, with special backgrounds such as biology, chemistry, economics, and engineering. They have been asked to examine the schedules of their specialty with particular reference to the two listings provided. Their task is to check the numerical listings with the special or abridged schedules and to select additional notations when expansion is considered necessary.

BULKHEADS	623.92
BULLHEADS	
BULRUSHES	582.542.4
BUNDLED CONDUCTORS	
BUOYANCY	532.3
BUOYS	627.9
BURNING	614.84
BURNING	632.187
BURNING	631.6.068
BURROWS	591.52
BUS (ELECTRICAL)	
BUTTERFLY VALVES	621.646.25
BUTTRESS DAMS	627.828
BYPASSES	
BYPRODUCTS	.002.67
CACTI	582.85
CADDISFLIES	595.74
CADMIUM	546.48
CADMIUM RADIOISOTOPES	
CAISSONS	624.157
CALCAREOUS SOILS	631.411.2
CALCITE	549.742.111
CALCIUM	546.41
CALCIUM CARBONATE	661.652.5
CALCIUM CHLORIDE	661.842.321
CALCIUM COMPOUNDS	631.416.7
CALCIUM COMPOUNDS	661.842
CALCIUM HYDROXIDE	
CALCIUM SULFATE	
CALIBRATIONS	53.089
CALICHE	661.471.3
CAMERAS	771.3
CAMP SITES	796.54.006
CAMPING	796.54
CANADA GOOSE	598.413
CANAL CONSTRUCTION	626.13
CANAL DESIGN	626.12
CANAL EMBANKMENTS	626.17
CANAL LININGS	626.3
CANAL SEEPAGE	626.826
CANAL SEEPAGE	626.14
CANALS	556.53
CANALS	626.1
CANNERIES	639.2.068
CANOPY	551.584.2
CANTILEVERS	624.35

Figure 2 . Alphabetic List.

*	624.15
*FOUNDATIONS	624.15
FLEXIBLE FOUNDATIONS	624.15
FOUNDATIONS	624.15
DAM FOUNDATIONS	624.152.63
SHEET PILING	624.152.63
MOVABLE DAMS	624.152.632.1
PILES (FOUNDATIONS)	624.154
TIMBER PILES	624.154.2
CONCRETE PILES	624.154.3
STEEL PILES	624.154.7
FRICTION PILES	624.155
PIPE DRIVING	624.155
CAISSONS	624.157
EARTHQUAKE ENGINEERING	624.159.1
REINFORCEMENT	624.159.4
ABUTMENTS	624.16
PIERS	624.16
*	624.19
*TUNNELS & TUNNELLING	624.19
PRESSURE TUNNELS	624.19
TUNNEL CONSTRUCTION	624.19
TUNNELING	624.19
TUNNELS	624.19
TUNNEL FAILURE	624.19.004.64
TUNNEL LININGS	624.19.02
TUNNELING MACHINES	624.19.02
TUNNEL DESIGN	624.191.1
TUNNEL HYDRAULICS	624.196
BRIDGE CONSTRUCTION	624.21
BRIDGES	624.21
BRIDGE DESIGN	624.21:72.01
CANTILEVERS	624.35
STRUCTURAL DESIGN	624.4
STRUCTURAL MODELS	624.9.001.57
ROOFS	624.91
*	625
*RAILWAY . HIGHWAY	625
AGGREGATES	625.07
GRADING	625.071
RAILROADS	625.1/.6
ROAD CONSTRUCTION	625.7
ROADS	625.71
ROADBANKS	625.714
ROAD DESIGN	625.72
CULVERTS	625.745
GULLIES	625.745
PAVING	625.8
ASPHALTIC CONCRETE	625.855.3
SOIL AGGREGATES	625.855.5

Figure 3. Notational List

Step 2: Cards for the notations to be inserted are being punched as before but without descriptor. These will then be sorted with the previously punched cards.

Step 3: A print of all UDC notations and their English language equivalent will be made to produce a subset of UDC water schedules. This will be done by matching each notation with the AIP/UDC tape through a computer program. For those notations in which no match occurs, a default will be made to print what occurs on the card containing the notation. In this way a description will be printed in the subset of schedules for all new or derived notations.

OBSERVATIONS AND RESULTS TO DATE

The numerical printout resulting from phases 1 and 2 confirmed that the classes most frequently used for water resources information are Classes 5 and 6 (Science and Technology). The ubiquity of water is obvious in that notations from all classes, except Class 2 (Religion), appeared. Only small parts fell into the socio-economic areas.

Less that 20% of the terms were not given a notation. Ambiguous terms such as "comparative benefits" or "limiting factors" fell into this category. The classifier must ask: "Comparative benefits in what context?" or "Limiting factors in what area?". In spite of the knowledge that such terms are related to water and are defined through the synonyms and related terms listed below each descriptor in the thesaurus, the conceptual area or areas into which these descriptors fall are either too numerous or too ambiguous to be given a specific UDC notation. On the other hand, one of the most important features of UDC is that it is an aspect classification, showing subjects in different contexts. In some cases several notations can be listed, e.g.

		Aspect
CONDENSATION	536.423	Physics(Heat. Thermodynamics.)
CONDENSATION	551.574	Meteorology.
CONDENSATION	556.131.3	Hydrologic Cycle.

At this stage the work has resulted in the development of an effective water thesaurus/concordance. Because we have not yet had time to produce and publish a Canadian thesaurus/concordance, a copy of the U.S. Thesaurus with UDC notations typed alongside the descriptors is

currently being used. (Figure 4) Additional vocabulary not found in the U.S. Thesaurus has been added as an addendum.

The Thesaurus/Concordance and the Special Water Schedules, to be completed in phase 3, will be issued in two volumes. These will be available in draft form after a testing period of indexing and searching in the WATDOC System. It is planned to have students use these tools on a trial basis during the summer.

IMPLEMENTATION IN WATDOC SYSTEM

Indexing of documents will not be restricted by the concordance. Indexers will be instructed to follow it whenever possible and to use the concordance as a primary tool for the selection of terms and subject classes. In an automated system some degree of consistency is imperative for the efficient retrieval of documents. Further, the schedules will be kept up to date through computer programs being developed for the production of the special water schedules. If efficient programs can be developed for such updating they can contribute to the eventual development of program packages for the automated updating and printing of schedules. As suggested by Wellisch (1971), the weakness of UDC in updating would be largely eliminated if compact editions could be generated from a central magnetic tape store.

Aquatic Life
Fish
Wildlife
RT Anadromous Fish
Animal Groupings
Eels
Fish Migration
Fish Types
Freshwater Fish
Wildlife

Catalysts 54-4
RT Chemcontrol 577.1
Chemical Reactions
Chemicals 66.097
Enzymes
Inhibitors 678.044
Retardants

Catastomids
USE Suckers

Catch Crops
BT Crops
RT Disasters
Drought Tolerance
Droughts
Dry Farming
Farm Management
Freezing
Land Use
Storms
Water Injury
Winterkilling

Catchment Basins
USE Watersheds (Basins)

Catfishes 597.554.4
UF Blindcats
Blue Catfish
Headwater Catfish
Ictalurus Catus
Ictalurus Furcatus
Ictalurus Lupus
Noturus Species
White Catfish
BT Aquatic Animals
Aquatic Life
Fish
Freshwater Fish
Pan Fish
Wildlife
NT Bullheads
Channel Catfish
Madtoms
RT Rough Fish

Cathodes 621.3.035.22
BT Electrodes
Equipment
NT Cathodic Protection
RT Corrosion Control
Electrochemistry
Electrolysis

Cathodic Protection 620.197.5
BT Cathodes
Electrodes
Equipment
RT Coatings
Corrosion
Linings
Pitting (Corrosion)

Cation Adsorption
BT Adsorption
Chemical Properties
Ions
Properties
RT Anion Adsorption
Cations
Clay Minerals
Ions

Cation Exchange
BT Chemical Reactions
Ion Exchange
Ions
Separation Techniques
Treatment
Water Treatment
RT Anion Adsorption
Anion Exchange
Cations
Control
Demineralization
Permselective Membranes
Quality Control
Water Quality
Water Quality Control
Water Treatment

Cations
BT Ions
RT Cation Adsorption
Cation Exchange
Electrolytes

Cattails 582.522.1
BT Amphibious Plants
Aquatic Life
Aquatic Plants
Monocots
Rooted Aquatic Plants
RT Aquatic Weeds
Riparian Plants

Cattle 636.2
BT Domestic Animals
Livestock
Mammals
Ruminants
RT Feed Lots

Cavefishes 597.24
BT Aquatic Animals
Aquatic Life
Fish
Freshwater Fish
Wildlife
RT Caves

Caverns
USE Caves

Caves 551.35.054.12
UF Caverns 551.442
BT Geology
Geomorphology
RT Cavefishes
Karst
Karst Hydrology
Limestones
Sinks
Travertine

Cavitation 532.528
BT Fluid Mechanics
Hydrology
RT Aeration
Bubbles
Corrosion
Erosion
Flow
Flow Separation
Flumes
Hydraulic Structures
Hydrofoils
Hydrologic Aspects
Impellers
Irrigation Water
Negative Pressure
Pressure
Reaeration
Scour
Stream Erosion
Vortices

Celerity
RT Rates
Standing Waves
Waves (Water)

Cells (Biological)
USE Cytological Studies

Cellulose 547.458.8
BT Carbohydrates 661.728
Organic Compounds
RT Fibers (Plant) 676.16
Lignins
Lumber
Pulp Wastes
Resins
Vascular Tissues
Wood Wastes

Cement Grouting 666.97.033.14
BT Grouting
RT Cements
Mortar

Cements 666.94
BT Construction Materials
NT Asbestos Cement 691.54
Portland Cements
RT Adhesives
Aggregates
Alkali-Aggregate Reactions
Asphalt
Cement Grouting
Clays
Concrete Additives
Concrete Mixes
Concrete Technology
Concretes
Diatomaceous Earth
Grouting
Linings
Masonry
Mortar
Paints
Pozzolans

Figure 4 . Example of Concordance

WATER THESAURUS/CONCORDANCE

REFERENCES

AITCHESON, Jean, 1970, "The Thesaurofacet: A Multipurpose Retrieval Language Tool", Journal of Documentation, v. 26, No. 3, p.187-203.

BRITISH STANDARDS INSTITUTION, 1963, Guide to the Universal Decimal Classification (UDC), British Standards House, London.

FOSKETT, A.C., 1969, The Subject Approach to Information, London, Bingley.

FREEMAN, R.R., ATHERTON, P., 1968, Final Report of the Research Project for the Evaluation of the UDC on the Indexing Language for a Mechanized Reference Retrieval System, Report AIP/UDC-9.

LLOYD, G.A., 1972, "The Universal Decimal Classification as a Switching Language", Subject Retrieval in the Seventies, School of Library and Information Services, University of Maryland, Greenwood, p. 116-125.

----------, 1972, Head, Classification Dept., Fédération Internationale de Documentation, The Hague, Personal Communication, 26 May, 1972.

MERCIER, Marcel A., 1971, "Study of UDC and other Indexing Languages through Computer Manipulation of Machine Readable Data Bases", University of Alberta, Edmonton, Master's Thesis.

MILLS, Jack, 1964, The Universal Decimal Classification, Rutgers Series on Systems for the Intellectual Organization of Information, v. 1, Graduate School of Library Service, Rutgers,New Brunswick, New Jersey.

OHMAN, E., OLIVECRONA, C., 1971, "Some Notational, Hierarchic and Syntactic Problems in connection with Concordance between UDC and Thesauri", Symposium on UDC in relation to other Indexing Languages, Herceg Novi, Yugoslavia.

PERREAULT, Jean M., 1969, An Introduction to UDC: A Programmed Text, London, Bingley.

RIGBY, M., 1970, "Computers and the UDC: A Decade of Progress", prepared for FID/CCC-M and 35th FID Congress, Buenos Aires.

SAVELJEVA, A.N., 1971, "On Revising the UDC Schedules, Scientific and Technical Information", Nauch-Tech. Inf. 2, U.S.S.R., Ser. 1, No. 8

SCRIBOR, E., 1971, "UDC as a Classification Connecting Various Information Languages", Current Problems in Information and Documentation, Central Institute of Scientific, Technical and Economic Information, May-June, Poland.

STUEART, R.D., 1971, "An Analysis of the UDC as a Term System for Nuclear Science and Technology", Library Resources and Technical Services, v. 15, no. 3, p. 399-411.

WELLISCH, H., 1970, The Universal Decimal Classification: A Programmed Instruction Course, School of Library and Information Services, University of Maryland.

----------, 1971, "Organisatorische Neurodunung des DK-Systems" (Organizational Renewal of the UDC System) Nachrichten für Dokumentation 22 nr. 2, p. 55-63.

----------, 1972, "A Flow Chart for Indexing with a Thesaurus", Journal of American Society for Information Science, v. 23, no. 3, p. 185-194.

WILSON, T.D., 1972, "The Work of the British Classification Research Group" Subject Retrieval in the Seventies, School of Library and Information Services, University of Maryland, Greenwood, p. 62-71.

WRIGHT, J.E., 1969, "Steps in Development of BS 1000", ASLIB Proceedings, v. 21, no. 5.

UN SYSTEME D'INFORMATION URBAIN ET REGIONAL, APPLIQUE A LA REGION DE L'ESTRIE, PROVINCE DE QUEBEC. (AN URBAN AND REGIONAL INFORMATION SYSTEM, AS APPLIED TO THE EASTERN TOWNSHIPS REGION, PROVINCE OF QUEBEC)

Pierre Lacasse
Centre de Recherche en Aménagement Régional,
Université de Sherbrooke, Sherbrooke, Qué.

RESUME

Nous traitons ici de la mise en application d'un système d'information urbain et régional dans la région de l'Estrie, province de Québec. Ce système qui consiste principalement en une base de données et une base géographique, a pour but de recueillir, d'emmagasiner et de contrôler un flot dynamique d'informations sur cette région: population, ressources, services ... pour fins de recherche, de planification, de développement et de gestion urbaine et régionale. (This paper illustrates the implementation of an Urban and Regional Information System in the Eastern Townships region of the province of Quebec. This system, which mainly consists of a data base and a geographical base, aims to manage a dynamic flow of data on this region: population, resources services ... for urban and regional planning and development purposes.)

Il existe au Canada de multiples projets d'échange d'information: centres d'échanges de données en sciences sociales, réseaux d'échange d'information urbaine, réseau de télécommunication entre universités ... Parallèlement se multiplient les systèmes d'information selon les divers champs d'intérêt.

Si nous restreignons notre point de vue au seul domaine des systèmes d'information basés sur une représentation géographique des données, on peut d'ores et déjà voir se dessiner au Canada des développements intéressants. C'est ainsi que se sont développés à titre d'exemples: le système de la ville de Vancouver; le SIUR de la ville de Québec; le GRDSR de Statistique Canada; le système de la commission de la capitale nationale; le CGIS (Canada Geographic Information Systems), partie de l'inventaire canadien des terres; la banque de données des environs de l'aéroport de Ste-Scholastique ...

Dans la perspective de contribuer d'une façon valable et continue

au développement de la région de l'Estrie, et d'apporter un support aux efforts de planification, de développement et d'aménagement poursuivis en cette région, l'Université de Sherbrooke, par son Centre de Recherche en Aménagement Régional, a décidé d'implanter un système d'information urbain et régional.

Dans les lignes qui vont suivre, nous traiterons premièrement des objectifs du système; deuxièmement de son organisation; troisièmement de ses principales composantes.

LES OBJECTIFS DU SYSTEME

Le système dont il est question ici, et qui se caractérise par un ensemble de fonctions gravitant autour d'une base de données et d'une base géographique, a pour but de conserver à jour les données sur l'état présent des ressources et des paramètres importants des ensembles administratifs de la région de l'Estrie. Ceci afin d'offrir un service de statistiques, de références et d'analyse le plus complet possible. Cette orientation découle directement des objectifs qui ont été fixés au départ, et que l'on retrouvera au tableau 1.

Quoique le système dans son application se penche sur les problèmes de l'Estrie, celui-ci peut s'adapter à d'autres ensembles et à des échelles différentes; c'est pourquoi les objectifs généraux énumérés ici ne sont pas définis en fonction d'une région particulière.

L'ORGANISATION DU SYSTEME

L'on voudra bien se reporter à la figure 1 (relations fonctionnelles du système) pour l'illustration de cette organisation.

Le système consiste en un ensemble d'éléments intégrés, et gérés d'une façon centrale par ordinateur, pouvant recueillir, emmagasiner, traiter et contrôler un flot dynamique d'informations touchant le milieu: la population, l'environnement et les ressources, l'utilisation du sol, les services et autres; afin de fournir un grand nombre de services aux divers usagers.

Le système se divise, pour fins d'organisation et d'administration en quatre grandes parties: (1) l'entrée, (2) la banque proprement dite, (3) la gérance et le traitement, (4) la sortie.

L'Entrée

UN SYSTÈME D'INFORMATION RÉGIONAL

TABLEAU I

LES OBJECTIFS DU SYSTÈME

- CUEILLIR ET TENIR À JOUR L'INFORMATION NÉCESSAIRE EN MATIÈRE URBAINE ET RÉGIONALE, SELON LA DIVERSITÉ DES SOURCES.
- STRUCTURER LES DONNÉES EN FONCTION DES BESOINS DES UTILISATEURS.
- TRAITER LES DONNÉES SOUS FORME D'APPLICATIONS DIVERSES, POUR RÉPONDRE AUX DEMANDES DES UTILISATEURS.
- OFFRIR UN SERVICE COMPLET DE STATISTIQUES, DE RÉFÉRENCES ET D'ANALYSE ET DE DIFFUSION DE L'INFORMATION DISPONIBLE.
- METTRE AU POINT DES TECHNIQUES NOUVELLES DANS LE DOMAINE DE L'INFORMATION URBAINE ET RÉGIONALE...

UN SYSTEME D'INFORMATION REGIONAL

Dans cette partie nous considérons premièrement la fonction d'acquisition des données, dont le rôle est de procéder à des cueillettes spécifiques, et de récupérer les données selon diverses sources; deuxièmement les principales sources d'accès aux données urbaines et régionales.

La Fonction Acquisition de Données (-4-)*. Cette fonction, dont le rôle principal est d'acheminer vers le système les données conformes aux objectifs fixés, consiste à analyser les besoins des utilisateurs; à recueillir, selon les sources et les modes appropriés, l'ensemble des variables à emmagasiner; à en préciser les mécanismes de mise à jour.

Pour remplir son rôle, cette fonction procède à des cueillettes spécifiques, à la récupération de données au niveau des opérations courantes des divers organismes oeuvrant dans la région, et fait appel aux sources les plus importantes qui possèdent des données sur la région.

Les Sources. Ces sources ne font pas partie, à proprement parler, du système, elles correspondent plutôt à l'environnement avec lequel le système (par la fonction acquisition de données) entend entrer en relation. Les principales sources à l'heure actuelle, qui sont de plusieurs ordres, sont énumérées ci-dessous.

Le centre de documentation (-1-) du Centre de Recherche en Aménagement Régional (CRAR), endroit où l'on regroupe toutes les publications et documents d'intérêt local et régional. Ce centre analyse la documentation et fournit à la banque un ensemble de références utiles aux organismes régionaux.

Les banques fonctionnelles (-2) de données régionales, avec lesquelles le système établit un lien. On peut citer comme exemples, les banques de données de la santé et de l'éducation, et autres banques fonctionnelles servant aux opérations courantes des divers organismes de la région.

Les sources extérieures (-3-) de données, qui sont ordinairement des organismes supra-régionaux (exemple le B.S.Q. et Statistique Canada), avec lesquels le système entend conserver des relations étroites et ainsi donner accès à ses usagers à un éventail plus complet d'informations; soit sous forme de données quantifiables, soit sous forme de références à des publications de ces mêmes organismes.

La Banque proprement dite

Celle-ci est constituée des trois composantes suivantes:

---

* Les chiffres entre tirets se rapportent à la figure 1.

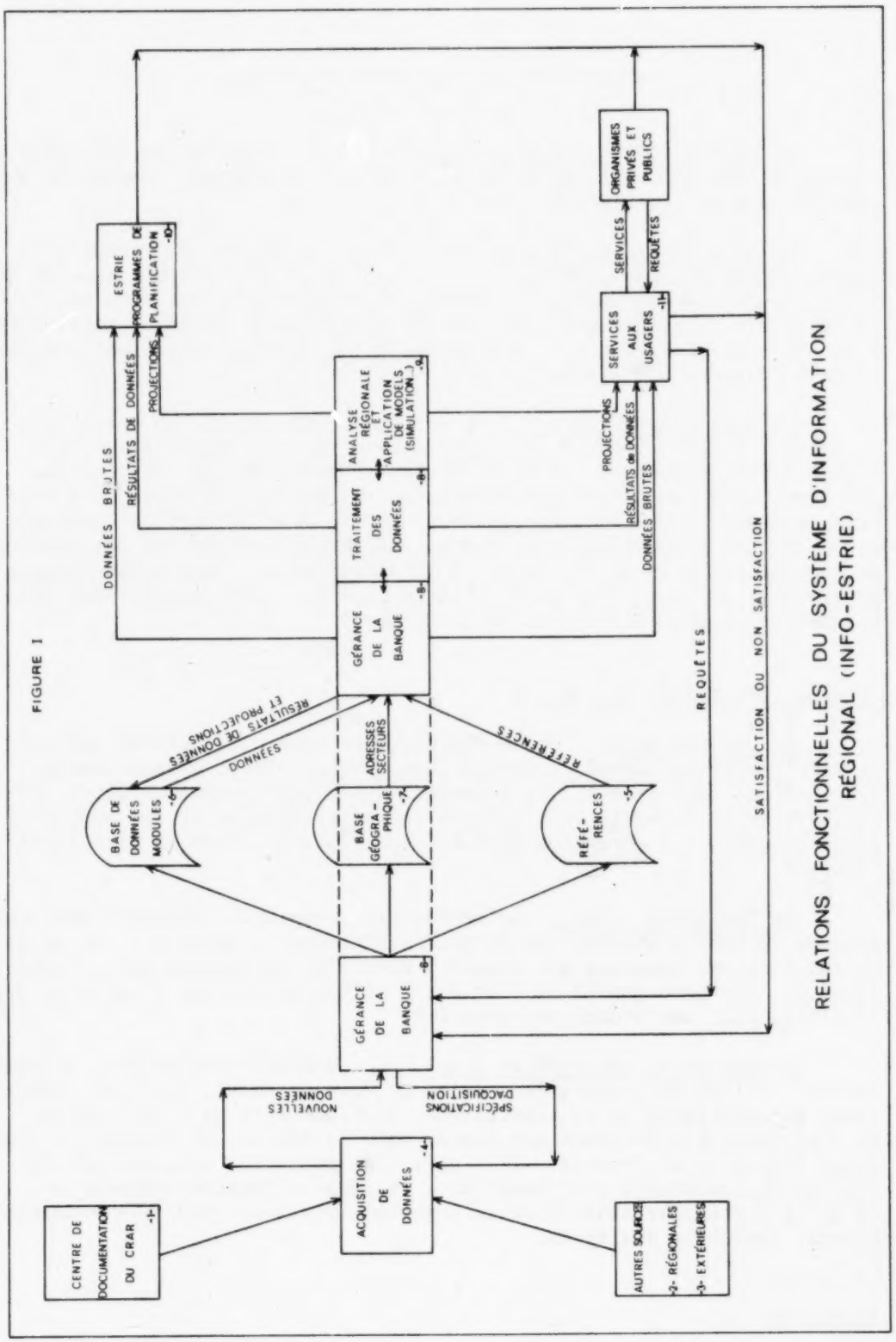


FIGURE 1

RELATIONS FONCTIONNELLES DU SYSTÈME D'INFORMATION RÉGIONAL (INFO-ESTRIE)

Le Module des Références (-5-). Ce module consiste en des index de références aux documents ou données, à caractère régional, conservés aux sources ci-haut mentionnées.

La Base de Données (-6-). Cette base se compose d'ensembles de données démographiques, économiques, sociologiques et autres, appelés modules, et de sous-ensembles de données se rapportant à ces modules, appelés fichiers. Ces données sont emmagasinées sur des supports appropriés à la suite de cueillettes et conservées à jour par le truchement des opérations courantes de certains organismes.

La Base Géographique (-7-). Par base géographique on entend ici un genre d'index servant à localiser sur le terrain chaque point d'emmagasinement de données. Nous avons mis au point une géocodification particulière à cet effet, qui fournit au système les principes à l'établissement de cette base géographique. Cette géocodification consiste en un système d'identification topologique, complété par une assignation de coordonnées métriques, servant à la localisation. Cette base géographique sert de lien entre toutes les données qui sont emmagasinées dans chacun des modules de données de la banque.

Gérance, Traitement et Analyse

La Gérance (-8-). Cette fonction, en regard des divers modules et fichiers de la banque de données sert à contrôler l'emmagasinement des données, leur mise à jour informatisée et leur retrait. Cette opération établit les relations voulues entre les modules et les supports géographiques. La gérance contrôle l'inter-action entre la machine et l'usager.

Le Traitement (-8-). Cette fonction comprend l'ensemble des programmes servant à établir les relations nécessaires entre les variables, en fonction des requêtes des usagers, pour fins de comparaison, de corrélation, de regroupement et tout autre calcul permettant d'apporter des résultats pour des études particulières.

L'Analyse et les Modèles (-9-). Ces analyses régionales, qui consistent en l'établissement de modèles de développement, à l'aide d'analyses mathématiques et de simulations, impliquent la mise en relation de l'ensemble des facteurs qui jouent dans la région, et d'analyser les inter-réactions de ceux-ci. Il s'agit d'un processus qui, quoique du domaine du traitement proprement dit, le dépasse considérablement en ce qu'il porte pour ainsi dire un jugement de valeur, plutôt que de simplement regrouper des faits.

La Sortie

Les Formes de Sortie. Normalement les extrants de la banque se présentent soit sous forme de données brutes (données extraites des fichiers sur demande, comme des tables chronologiques), soit sous forme de résultats de données ou de projections (suite à des analyses spécifiques). Une forme ou l'autre venant alimenter d'une part des programmes spécifiques de planification régionale (-10), et d'autre part le service aux usagers, dont il est question ci-après.

Le Service aux Usagers (-11). C'est l'interface entre le système et l'extérieur ou l'environnement. Cette fonction consiste à établir les échanges entre la banque et les usagers, selon les divers types d'applications demandées. Ce service répond à des requêtes d'individus et d'organismes, privés ou publics pour fins de planification, de développement, d'aménagement et de gestion, selon leurs intérêts particuliers.

C'est également ce service qui sert de canal de "feed back" de la satisfaction ou de la non satisfaction de l'environnement en fonction des besoins. Donc, en plus d'avoir à informer du contenu de la base, d'analyser les requêtes, il a pour mission de faire des sondages vis-à-vis des nouveaux besoins.

LES PRINCIPALES COMPOSANTES DU SYSTEME

Parmi les principales composantes du système figurent ce que l'on pourrait considérer comme le noyau central ou la banque de données elle-même. On y fait référence en parlant de la base de données d'une part, et la base géographique d'autre part.

La Base de Données

Les données sont regroupées en une base intégrée, pour faire opposition à des assortiments (data sets), organisés spécifiquement pour une application particulière. Cette base est formée d'une série de modules ou groupes homogènes de données (exemple Environnement/ressources), lesquels modules se divisent en fichiers et sous-fichiers.

Pour fins d'entrée, d'emmagasinement et de réemmagasinement après agrégation ou autres traitements, les données sont regroupées selon divers niveaux d'intérêt; chaque niveau donnant lieu à la création d'un sous-fichier particulier.

Ces principaux niveaux sont pour le moment: le lot, la façade, l'îlot, le secteur de dénombrement, le secteur de recensement, la municipalité, le comté et la région. Mais il n'est pas exclu de déve-

UN SYSTÈME D'INFORMATION RÉGIONAL

TABLEAU II

LES PRINCIPAUX MODULES DE DONNÉES

- **INFRASTRUCTURE URBAINE ET RÉGIONALE**

  Fichiers: habitation et logement; transports et communications; équipements et services; propriétés; utilisation du sol...

- **SOCIAL**

  Fichiers: population; santé; bien-être; éducation; loisirs...

- **ACTIVITE ÉCONOMIQUE**

  Fichiers: main-d'oeuvre; industrie; commerce; économie régionale...

- **ENVIRONNEMENT / RESSOURCES**

  Fichiers: agriculture; forestrie; récréation; mines; eau; sol...

lopper d'autres niveaux, selon les intérêts; ainsi pourront se dessiner d'autres cellules de base servant à l'emmagasinement de données de diverses natures. Par exemple, on peut imaginer des zones écologique spécifiques à l'emmagasinement de données d'environnement; ou encore à des aires correspondant à des plans de zonage.

On retrouvera au tableau 2 les premiers modules préconisés par le système Info-Estrie, ainsi que des exemples de fichiers; les sous-fichiers ne figurent pas ici, mais comme il est fait mention plus haut ils correspondent à des niveaux.

La Base Géographique

Celle-ci est constituée par un ensemble de fichiers, résultats de l'inter-connexion des divers éléments de la réalité spatiale. Nous analyserons donc d'abord cette réalité spatiale, après quoi nous décrirons brièvement l'organisation elle-même de la base en fichiers.

La Réalité Spatiale. Si l'on analyse les structures naturelles et artificielles de l'environnement, on se rend compte qu'il existe trois composantes interreliées correspondant à la réalité spatiale ou administrative (selon le cas) où se déroulent les activités d'une société donnée (activités sociales, économiques, culturelles, récréatives) et où se situent les équipements, les services, les structures implantées par l'homme, et les matières premières qui supportent ces mêmes activités. Ces trois composantes sont: les réseaux (exemples: réseau routier, réseau hydrographique); les surfaces ou assises (exemples: lotissement ou propriétés, cellules écologiques); les limites administratives ou autres qui viennent circonscrire des champs d'intérêt et d'activités sectorielles (exemples: limites municipales, secteurs de recensement, zones scolaires). La figure 2 montre un exemple de cette organisation spatiale, que l'on retrouve en milieu urbain.

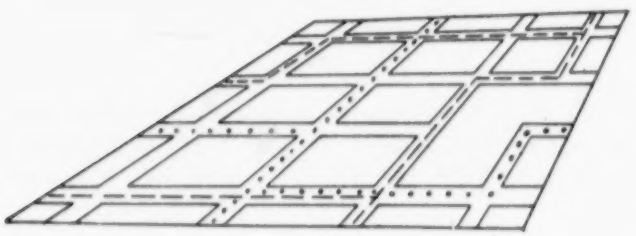
Chacune de ces composantes comporte des caractéristiques propres qu'il est bon d'expliciter.

Les réseaux sont ces ensembles de traits qui constituent l'infrastructure urbaine et rurale, de même que toutes autres structures linéaires naturelles. Comme exemple, on peut parler des réseaux routiers, des réseaux hydrographiques.

Les structures de surfaces de leur côté sont toutes les cellules de base où se situent les équipements et les ressources, et autour desquelles gravitent les activités d'un milieu donné. Exemples, les lots ou propriétés, les surfaces écologiques, les surfaces géomorphologiques.

FIGURE II

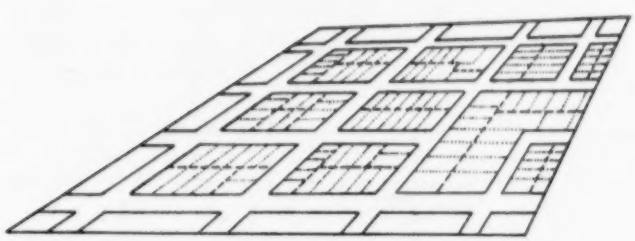
Les composantes de L'ORGANISATION SPATIALE URBAINE



LIMITES ADMINISTRATIVES

------ secteurs de recensement.

------ arrondissements municipaux.

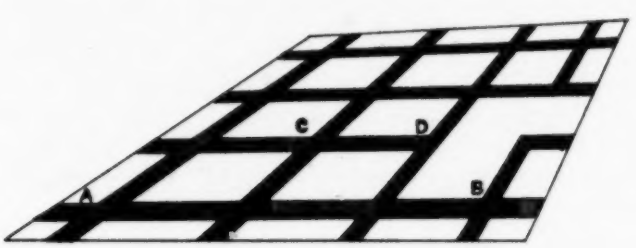


STRUCTURE DU LOTISSEMENT

——— ilôts

------ Façades d'ilôts

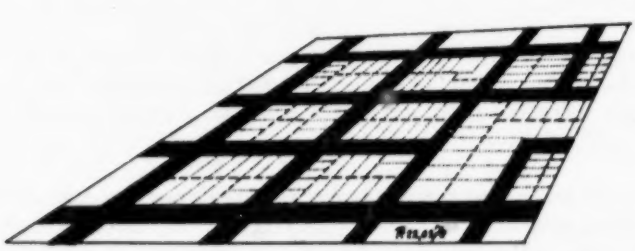
..... lots
(propriétés foncières)



RESEAUX DE RUES

AB: rue

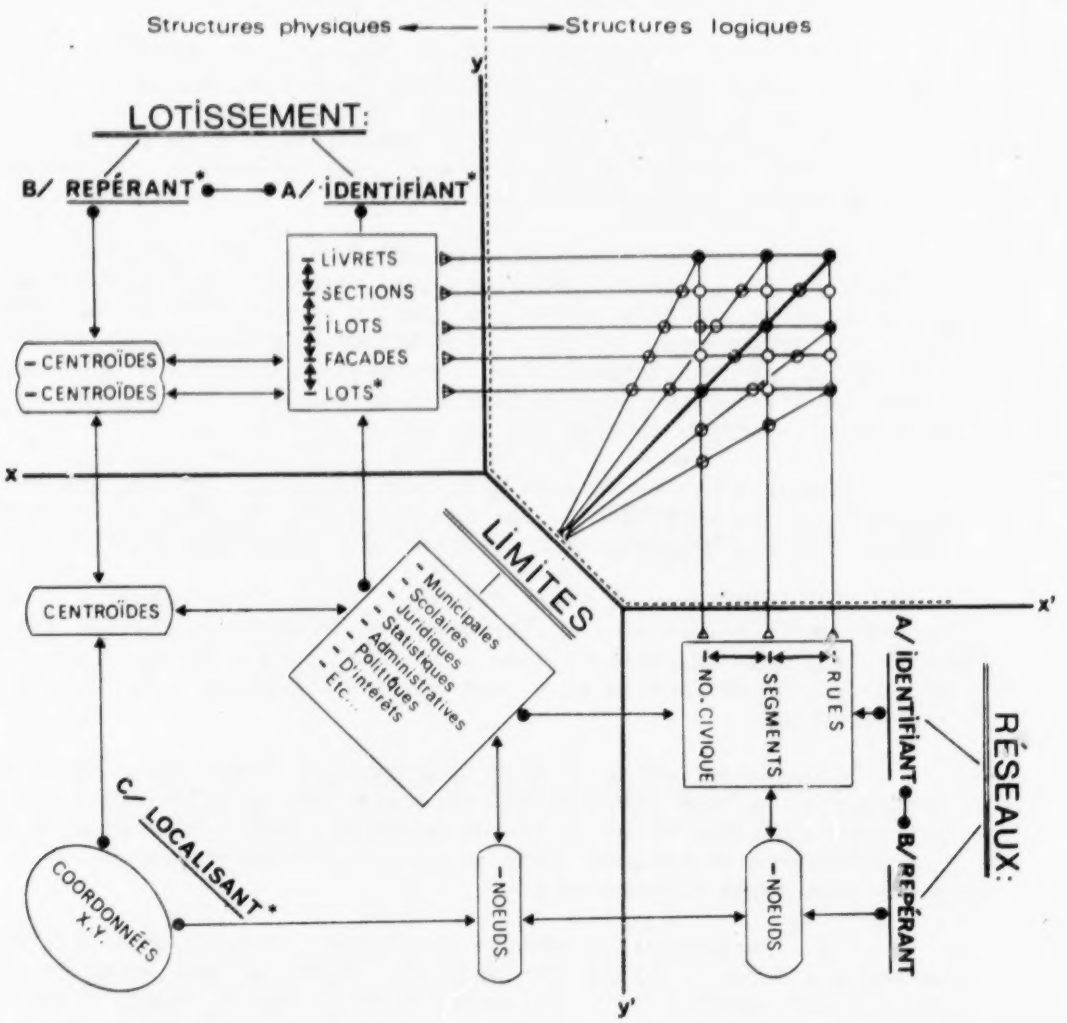
CD: segment de rue



STRUCTURE SPATIALE URBAINE

FIGURE III

EXEMPLE des ÉLÉMENTS de la RÉALITE SPATIALE URBAINE
et de leurs INTER·RELATIONS



IDENTIFIANT*: **Identification topologique.**
REPERANT* : **Agent de repérage.**
LOCALISANT* : **Localisation métrique.**
LOT* : Propriété foncière / ou assise d'aménagement.

B. Rivard, Info-ESTRIE, CRAR, Nov. 1972.

Ces cellules primaires servent à fournir un élément de base à tous les regroupements futurs, selon les champs d'intérêt.

Les limites correspondent à ces ensembles ou secteurs d'activités, à ces zones définies arbitrairement venant se superposer aux cellules de base et venant les regrouper, soit pour des fins de compilation statistique, soit pour des fins administratives ou autres. On pense ici aux limites municipales, aux zones scolaires, aux secteurs de recensement, aux secteurs agricoles ...

Chacune de ces composantes ou structures sont hiérarchisées et identifiées d'une façon topologique, par ensembles et sous-ensembles, et ainsi se définissent des niveaux d'agrégation possibles et de relations entre les structures. La figure 3 nous donne un exemple de cette identification, et de l'interrelation entre les éléments des diverses structures, en l'occurence la structure hiérarchique du lotissement et la structure du réseau routier.

Chaque élément, en plus d'un identificateur topologique est représenté par un centroïde ou par des noeuds de contours ou d'intersection de traits, auxquels sont assignées des coordonnées métriques.

Ces structures et leurs éléments, leurs interrelations à divers niveaux, ces limites et les regroupements de surfaces suivant les besoins d'agrégation, donnent lieu à la constitution de multiples fichiers et sous-fichiers dans la base géographique, fichiers dont il sera maintenant question.

L'Organisation de la Base Géographique. Comme nous venons de le mentionner, la base géographique est constituée de fichiers issus des principaux niveaux d'interrelation entre les hiérarchies de surfaces et les structures de réseaux d'une part, et de la superposition de limites à ces structures d'autre part.

Si l'on poursuit l'exemple de l'espace urbain, on voit l'utilité de définir trois fichiers décrivant les interrelations entre la hiérarchie du lotissement et le réseau des rues, et explicitant le jeu des limites superposées à ces structures; ce sont les fichiers au niveau du lot, au niveau des segments de rue et au niveau des limites. Ces trois fichiers suffisent à former toutes les combinaisons nécessaires au retrait des données de propriété. (Voir la figure 4, pour une représentation de ces fichiers, définis en fonction de l'espace urbain). Nous décrivons maintenant chacun de ces trois fichiers.

Le premier fichier, qui décrit le niveau du lot, comprend: le numéro du lot, les coordonnées du centroïde, le numéro civique inférieur situé sur ce lot, le numéro de la rue, une série de pointeurs servant

FIGURE IV

EXEMPLE DE FICHIERS DE LA BASE GÉOGRAPHIQUE, DÉFINIS EN FONCTION DE L'ESPACE URBAIN.

FICHIER AU NIVEAU DU LOT:

NO du LOT	CENTROIDE		NO CIVIQUE INF.	NO RUE	POINTEUR	POINTEUR	POINTEUR
	X	Y					

FICHIER AU NIVEAU DES SEGMENTS DE RUE:

NO RUE	NO SEG.	NOEUD INF.	COORD.		NOEUD SUP	COORD.		FAÇADE PAIRE	NO CIVIQUE		POINTEUR	FAÇADE IMPAIRE	NO CIVIQUE		POINTEUR
			X	Y		X	Y		SUP.	INF.			SUP.	INF.	

FICHIER DES SECTEURS:

1er CODE	NO du SECTEUR	2° CODE	DÉBUT	POINTEUR 1	FIN	POINTEUR 2

0 MUNICIPALITÉ
1 RECENSEMENT
2 ARRONDISSEMENT

0 LOT
1 FAÇADE
2 ILÔT
3 COORDONNÉES

à relier les lots entre eux et à relier les fichiers entre eux.

Le second fichier représentant le niveau des segments de rue, renferme entre autres choses tous les éléments nécessaires au retraçage automatique. Voici son contenu: un numéro de rue (correspondant à un nom, non exprimé ici); ce numéro correspond à l'identificateur topologique de la rue. Un numéro de segment, sous-partie de la codification de la rue (un segment étant défini comme une portion de trait entre deux intersections). Le numéro des noeuds inférieur et supérieur d'un segment (un noeud étant un point situé aux intersections de traits). Les coordonnées des noeuds inférieur et supérieur. Le numéro de l'îlot contenant la façade paire. Le numéro de l'îlot contenant la façade impaire. Les numéros civiques inférieur et supérieur de la façade. Enfin des pointeurs référant aux éléments de la façade.

Le troisième fichier, celui des secteurs, déterminé par les diverses limites qui viennent se surimposer aux structures, comprend de son côté: un premier code qui indique le type de secteur (ex. municipalité, secteur de recensement, arrondissement municipal); vient ensuite le numéro ou l'identificateur du secteur en question. Un deuxième code venant donner la façon dont ce secteur sera décrit (ex. par lot, par façade, par coordonnées). L'entrée "début", indiquant le premier élément d'une suite décrivant le secteur. L'entrée "fin", indiquant le dernier élément d'une suite descriptive. Les pointeurs référant aux premier et dernier éléments de la suite.

Ce fichier des secteurs sert spécialement à regrouper des surfaces (exemple des lots), pour des fins d'agrégations spécifiques de données. Ces secteurs peuvent par ailleurs correspondre à un palier ou niveau d'emmagasinement des données dans la banque, par exemple, toujours en fonction de l'espace urbain, une façade d'îlot, un secteur de dénombrement.

Un grand avantage de ce fichier des secteurs est de permettre d'éviter la répétition de codes d'appartenance à plusieurs ensembles, à chacun des records d'emmagasinement de données. Par ce truchement, on fait référence au contenu par une définition externe soit à l'aide des ensembles topologiques, soit à l'aide des coordonnées; tout cela sans perturber la codification des unités ou records en cause.

CONCLUSION

Après un an et demi de fonctionnement, Info-Estrie peut offrir

une conception tout à fait originale de développement d'un système de données à références spatiales; et les avantages qu'on peut déjà entrevoir pour la région touchent l'amélioration des systèmes administratifs tout autant que l'aménagement régional, l'organisation et la planification des ressources. On pense ici au domaine de l'évaluation, celui des services techniques et d'urbanisme; on se réfère également aux statistiques de population que l'on retrouve à divers niveaux d'agrégation, et qui peuvent servir à des études de projection et de corrélation de toutes sortes.

Parmi les réalisations que le système a pu concrétiser jusqu'à maintenant, on note:

- Du traitement des données de recensement à tous les niveaux, pour l'ensemble de la région.

- Du traitement en vue d'études sectorielles: tracé optimum pour transport scolaire; analyse des ressources humaines en loisir; élaboration d'un système d'identification de la propriété pour fins d'évaluation; compilation des sondages géologiques de la région et analyse géomorphologique; analyse de bassins hydrographiques; analyse d'accidents de circulation.

- Des représentations graphiques: tracés de réseaux municipaux et régionaux; représentation de traitements des données de recensement, par secteurs de recensement, par municipalités, pour la région; représentation d'accidents de circulation; représentation du pavage des rues.

- Des analyses SYMAP: répartition de la population de la région; densité de la population; analyse de la structure d'âges.

- L'implantation d'un système de références bibliographiques informatisées, sur la région de l'Estrie.

Pour ses différents travaux, Info-Estrie dispose de l'ordinateur de l'Université (IBM 360, modèle 40); un nouveau modèle est attendu pour septembre 1973 (IBM 370, modèle 145). Le système graphique PACER 930 (EAI), du laboratoire de cartographie, est également à sa disposition.

Le système Info-Estrie, tel que conçu, n'est pas spécifique à une aire géographique donnée, mais il a l'avantage de pouvoir être appliqué dans d'autres régions, et nous prévoyons par ailleurs l'intégrer à un réseau plus vaste d'échanges d'informations urbaines devant être implanté au Canada.

SCIENCE POLICY AND INFORMATION NETWORKS AN INTRODUCTION (POLITIQUE SCIENTIFIQUE ET RESEAUX DOCUMENTAIRES - UNE INTRODUCTION)

Jack E. Brown
National Science Library
National Research Council of Canada
Ottawa, Ontario K1A OS2

ABSTRACT

In 1970 Canada defined and began implementation of a science information policy designed to expedite the flow of S & TI. The NRC was assigned the responsibility for the development of a national decentralized S & TI system. The concept was not new but the directive provided formal recognition of and authority for the further development of the existing national system wherein the NSL serves as the focal point. Good progress has been made in speeding up responses to requests for information, in promoting and publicizing existing information services, and in developing a national SDI service. However, the geographical and political nature of Canada place almost insuperable constraints on any attempts to accelerate the rate of progress. A small planning group appointed by the President of the NRC has made recommendations, which if implemented could overcome these constraints and provide the impetus we now seek. (En 1970 le Canada a défini et commencé à mettre en oeuvre une politique de l'information scientifique conçue pour accélérer la circulation de l'IST. Le CNRC a été chargé de mettre au point un système national décentralisé d'IST. L'idée n'était pas neuve mais cette directive reconnaissait et autorisait officiellement le développement plus ample du système national existant, dont la BSN est la cheville ouvrière. Entre autres progrès importants, on a accéléré les réponses aux demandes de renseignements, favorisé et mieux fait connaître les services documentaires existants et perfectionné le service national de diffusion sélective de l'information. Toutefois, les conditions géographiques et politiques du Canada imposent à toute tentative d'accélérer le progrès des restrictions presque insurmontables. Une petite équipe de planification nommée par le président du CNRC a présenté des recommandations qui, si elles sont mises en oeuvre, pourraient triompher de ces obstacles et assurer la viguer tant souhaitée.)

SCIENCE POLICY

Two years ago, while presiding at an international conference on governmental responsibility for providing information to industry, the Right Honourable Wedgwood-Benn, then Minister of Technology for the UK, said that "in the 70s information of all kinds will emerge as the most valuable of our national and international resources".

I do not think Mr. Wedgwood-Benn was attempting to act as a prophet - he was simply drawing attention to the fact that there is a vital relationship between the industrial, economic, and social development of a country and the ability of that country to channel relevant information to the right person at the right time. With the acceptance of this axiom, the major countries of the world now regard the development of policies and techniques to ensure maximum exploitation of information, particularly scientific and technical information, as a matter of major concern.

Canada, in spite of our predilection for endless self-examination and procrastination, has taken positive action to achieve this maximum exploitation of information. Indeed, Canada appears to be the first country to have established a science information policy and to have taken steps to implement it.

In January, 1970 following a series of studies on scientific and technical information (S & TI) in Canada which culminated in the publication of Science Council Report No. 6, the NRC was assigned by the federal government responsibility for the development of a national S & TI system. The directive specified that the system should be developed "in concert with existing information organizations", should be under the general direction of the National Librarian, and that the National Research Council of Canada (NRC) should appoint an advisory board to formulate general policies.

At first glance this directive appeared to offer a completely new concept and approach which would facilitate the dissemination of information in Canada. However, this concept was not new, for over the years Canada had developed the foundations of a national S & TI system. Nevertheless it did provide formal recognition of and authority for the further development of an existing and successful operation.

By 1970 there were already two identifiable and viable networks within this national system. First, there is a national network of 270 of Canada's major libraries which through a variety of voluntary cooperative measures make their resources available nationally by means of loans and photocopies. The files of scientific and technical journals held by these libraries are linked by means of the computer-based Union List of Scientific Serials in Canadian Libraries. Since journals constitute at least 80% of the resources of a scientific and technical library, this network ensures that Canadians have access to 80% of the world's scien-

tific and technical literature. Here quite clearly the decentralized concept of utilizing resources wherever they exist is in full force.

The second network within this national system is the CAN/SDI program, a current awareness service wherein 1400 subscribers are alerted to the existence of recent papers covering their specific fields of interest as published in the world's scientific and technical literature. This too is a decentralized system with 400 subject or information specialists, located in all parts of Canada, serving as interfaces between the CAN/SDI system and the ultimate user.

The focal point of this national S & TI system is the National Science Library (NSL) or perhaps as it could be more correctly entitled, the Canadian Centre for S & TI. The NSL's resources and services, which have been developed in close cooperation with other major libraries (particularly federal, university, and industrial libraries) are designed to supplement and complement local and regional resources. This back-stopping operation is also intended to ensure that the scientific, engineering, and industrial communities of Canada have ready access to the information required in day-to-day activities. The weaknesses of the system - a lack of knowledge of existing services, delays in the delivery of needed documents or information, failure to recognize, strengthen and tap local sources of expertise, high communication costs - are inherent in the geographical and political nature of Canada. During the last two years our major concern has been to seek ways and means by which these weaknesses and constraints to progress can be overcome.

The directive to the NRC further stated that the development of a national S & TI system should be under the general direction of the National Librarian. Here again was an attempt to formalize and strengthen an existing situation. The National Library (NL) and the NSL have a long tradition of cooperation and collaboration in the development of information resources and services. In recent years, with the establishment of the National Library's responsibility for developing bibliographical standards, and with the increasing use by the NSL of mechanized techniques for processing information, the NL and NSL have established a variety of joint working groups or task forces to guarantee the establishment of compatible information handling systems regardless of the subject matter covered.

The need for a committee or group to advise NRC in dealing with national S & TI problems was recognized as early as 1958, when the Associate Committee on Scientific Information was established. The appointment of the Advisory Board for Scientific and Technological Information (ABSTI), in keeping with the third point in the directive to NRC, has created a body with the power and prestige to spark action which was not possible under the former Associate Committee.

SCIENCE POLICY

With the assignment to the NRC of new responsibilities for national S & TI services, considerable progress has been made in expediting the flow of S & TI in Canada. An increasing number of users of information are now familiar with services available to them; requests for publications and information are met more promptly (36 hour response time by the NSL); and through CAN/SDI 3800 researchers in government, industry, and universities are regularly alerted to current developments in their fields of interest. Despite these developments, it is very evident that a large percentage of the potential users of S & TI do not have ready access to information which is relevant and timely.

In May 1972 the President of the NRC appointed a small planning group to critically examine the current situation and to recommend steps which should be taken to accelerate the development of a national S & TI system. The group immediately recognized that we were endeavouring to develop not so much a national S & TI system, but rather a national network of scientific and technical information services. In addition to the constraints mentioned above, the group also identified factors which although less obvious, do constitute major roadblocks in the path of progress. As a result of its study, the group has made several recommendations relating primarily to the development of greater collaboration between the major information disseminating agencies in Canada, the identification of user needs, and the utilization of specialized information centres as nodes within the national system. These recommendations, if implemented, could provide the impetus to progress which we are now seeking.

The papers to be presented at this afternoon's session relate directly to the problems I have been discussing. Three of the papers describe ongoing specialized information systems which could logically constitute nodes within the national system. The fourth paper, by our keynote speaker, Dr. Aurèle Beaulnes of the Ministry of State for Science and Technology, deals with the problems of defining a science policy and the role of science information. MOSST is vitally concerned with having available all relevant information which will assist the Department in making decisions regarding the promotion of scientific and industrial research activities in Canada. In this connection, the NSL is working closely with MOSST to establish an inventory of scientific activities in Canada. MOSST's expertise and guidance is also invaluable to the NRC/NSL in creating a closer liaison between federal and provincial agencies having major responsibilities for the dissemination of scientific and technical information.

EVOLUTION OF CANADIAN SYSTEM FOR GEOSCIENCE DATA, 1964-73 (EVOLUTION DU SYSTÈME CANADIEN DE DONNEES GEOSCIENTIFIQUES, 1964-73)

Cornelius F. Burk, Jr.
Canadian Centre for Geoscience Data
Ottawa, Ontario K1A 0E8

ABSTRACT

A programme to apply modern information technology to the communication of geological data at the national level was begun in 1964. During the past nine years the concept of the Canadian System for Geoscience Data has evolved in response to advancing technology, and provides an illustration of how such technology can influence the conduct of science. In 1964 the Canadian System concept was predicated on standardized codes and formats for data; in 1967 standardized intellectual content was regarded as crucial; today, I consider systems intercommunication techniques as the key ingredient for a successful national system. Feedback to users from such intercommunication will promote a higher degree of standardization as a natural consequence. Thus, standards for data will become a product of information technology, not a prerequisite to using it. (Un programme visant à appliquer les techniques d'information les plus modernes à la communication de données géologiques au niveau national, fut mis à exécution en 1964. Durant les neuf dernières années, le concept du Système Canadien de Données Géoscientifiques a évolué devant le progrès de la technologie, et nous fournit une illustration de la façon dont cette technologie peut influencer la conduite de la science. En 1964, le concept du Système Canadien fut basé sur l'utilisation de codes et formats standards pour les données; en 1967, des "standards de contenu intellectuel" étaient considérés comme essentiels; aujourd'hui, je pense que les techniques d'intercommunication entre systèmes constituent l'élément-clé pour l'obtentien d'un système national efficace. Cette intercommunication aura pour effet, chez les usagers, de promouvoir un plus haut degré de standardisation. Alors, les standards pour les données seront le produit des techniques d'information, et non pas un prérequis à leur utilisation.)

GEOSCIENCE DATA

INTRODUCTION

The value of basic observations and measurements in the earth sciences is especially high in Canada, since in addition to playing their obvious role in the conduct of scientific research and development, geoscience data are among the *sine qua non* for development of our mineral and fuel resources and, increasingly, for the preservation of our environment. The current concern for energy sources, especially oil and gas, is but one example of an activity of deep economic and social concern in which geoscience data play a fundamental role.

In attempting to apply geoscience data more effectively to the solution of such broad problems, and to help advance the science in general, a small group of Canadian geologists began in 1964 to explore ways in which modern information technology could be applied on a national scale to the storage, retrieval and communication of geoscience data. Now, nine years later, it is instructive to review the developments that have transpired and to note, above all, how the availability of new technology has influenced proposals and activities related to the exchange of scientific data, and potentially to the conduct of science itself.

I wish to make it clear that I am using the much abused term "data" in the specific sense of basic scientific observations and measurements, that is, the records of what geoscientists or their instruments have seen or measured at some particular place and at some particular time. Such "data", with reduction and other processing, grade into higher levels of information including conclusions, syntheses and opinions. Although there is no objective way of distinctly dividing these various categories, the identification of the end-members in the context of a given study or investigation is usually clear. It is the "data" end-member of the information spectrum to which the Canadian System for Geoscience Data has addressed itself.

ORIGIN OF THE CONCEPT

At the 16th annual meeting of the National Advisory Committee on Research in the Geological Sciences, on 27 April 1964, K. R. Dawson of the Geological Survey of Canada described the contents and requirements of a proposed computer-based file for analytical and other data derived from laboratories of the Geological Survey. His presentation provoked a detailed discussion which included a proposal by A. D. Baillie, British American Oil Co. Ltd. (now Gulf Oil Canada Limited),

> "...that a subcommittee should be set up to make a survey of computer equipment and attempt to formulate a workable system of data compilation and retrieval that might be used not only by the Geological Survey, but by organizations across the country."

This suggestion marked the beginning of a series of events which lead ultimately to the Canadian System for Geoscience Data. Thus, in late 1964 the concept of a "national system" and the feasibility of developing it were examined in a special study by S. C. Robinson of the Geological Survey of Canada. His unpublished report (Robinson, 1965) advocated the establishment of a national system based primarily on the use of 80-column punched cards for data storage and the establishment of a parallel national index to these data. Each data file would be controlled by its owner, but would include national standards for the data, including standardized reference numbers, a method of recording geographic location, and a standard classification and a method of description for rocks and other geological materials. Although third-generation computers were becoming available at this time, the rationale for these proposed standards was based on the capabilities of unit-record (card-sorting) machines, since this was the only type of equipment widely accessible to geologists.

CONCEPT DEFINITION

Following the feasibility study, the National Advisory Committee on Research in the Geological Sciences established an *ad hoc* committee under the chairmanship of S. C. Robinson to examine in detail the concept of a "national system" for storage and retrieval of geological data and to recommend on the means for implementation. A two-year study involving over 40 geologists and other scientists resulted in publication of interim (Robinson, 1966) and final (Brisbin and Ediger, 1967) reports which defined the National System and presented the results of pilot studies to illustrate how the concept might be applied in practice.

The concept which emerged envisaged an owner-controlled network of computer-based files, "linked" by the common use of various standards for geological data recorded therein. In addition, a computer-based national index to the sources of all data on Canada, computer-processable or otherwise, was included. Stress was placed on the decentralized, independant control of files in the network, and on the fundamental need for geologists to establish standards for their data. The proposed organization of the National System, as published in the 1967 report, is reproduced in Figure 1.

Use of the term "system" in this way was misleading to those who generally equate "system" with a tightly circumscribed set of procedures, usually operating under one roof. As presented in 1967, the National System could perhaps have been described more accurately as a "National Programme for development of compatible computer-based data files". Nevertheless, from a broader point of view, the concept did indeed constitute a system (cf. Canadian Postal System).

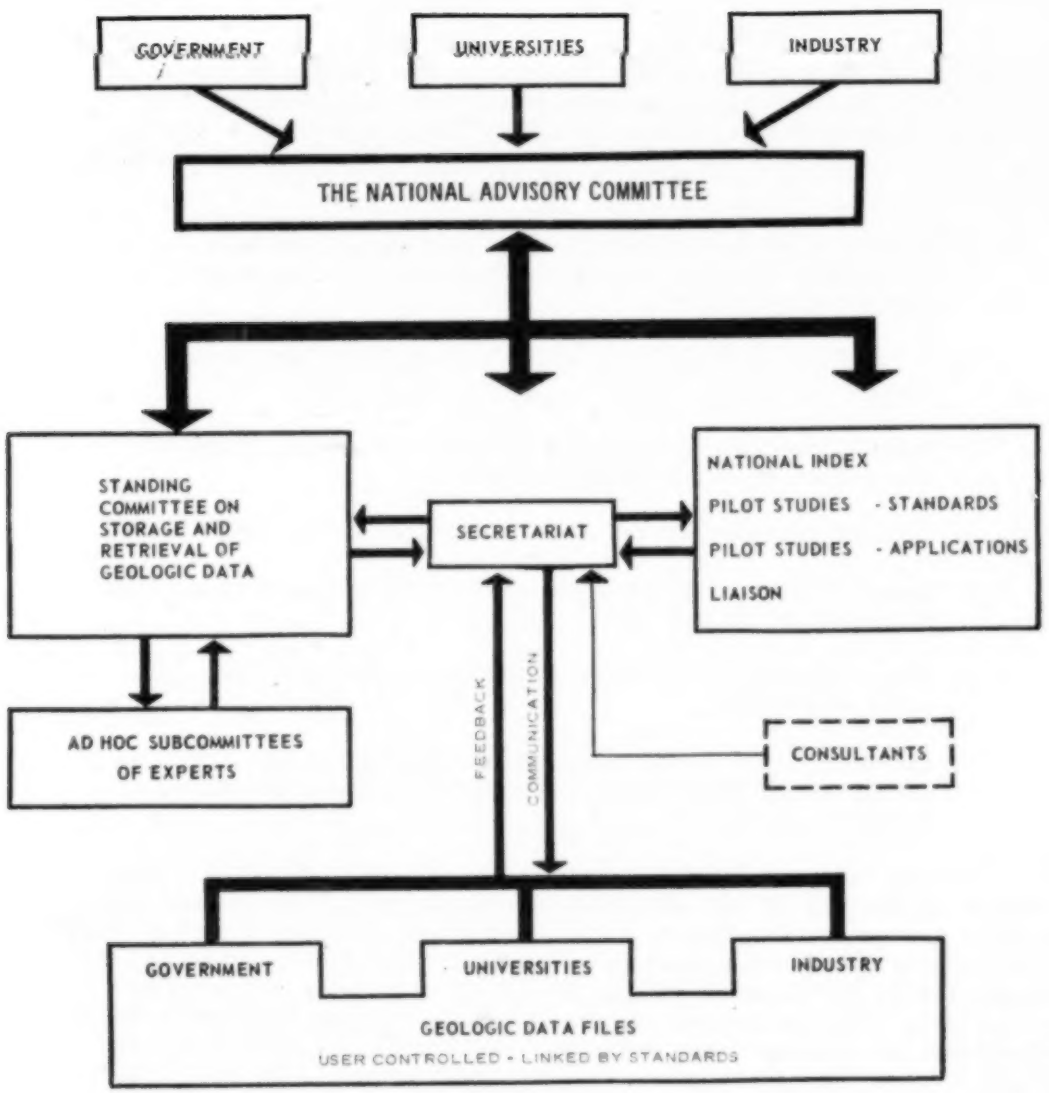


Figure 1. Organization of national system for geological data as proposed in 1967 by National Advisory Committee on Research in the Geological Sciences (reproduced from Brisbin and Ediger, 1967, Fig. 4).

IMPLEMENTATION

Following discussion with various national bodies, especially the Provincial Ministers of Mines Conference and the Science Council of Canada, the Geological Survey of Canada began implementing the various recommendations of the National Advisory Committee's 1967 report for establishment of a National System. Among the major events were:

1. Permanent advisory committee established to oversee implementation i.e. NACRGS Subcommittee on Computer Applications (1967).

2. Pilot projects begun on computer-based national index (1967).

3. Special research grants awarded for developing computer-based data files in geology (1968).

4. Interim Secretariat for Geoscience Data established (1968).

5. Canadian Centre for Geoscience Data established to replace Secretariat (1970).

6. First edition of *Canadian Index to Geoscience Data* published (1970).

7. Advisory Board to Canadian Centre for Geoscience Data appointed by Department of Energy, Mines and Resources (1971).

During the period 1967-70, three major lines of activity were followed in pursuit of the National System concept; then renamed the Canadian System for Geoscience Data: Compilation of the *Canadian Index to Geoscience Data*, using systems software contributed by Imperial Oil Limited (SIS); development by committees of techniques and standards for building files of geological field data and mineral deposits data; and university research into data-base management and data files of various types. The development of standards for data was the primary goal behind most of this work, as had been recommended by the *ad hoc* committee. Although not overtly expressed at the time, the underlying reason for this goal - with emphasis on standards for data content, rather than codes and formats - was the assumption that available computer technology could not be used for the national communication of data on any significant scale without such standards. Of course there was also the desire to promote standardization on purely scientific grounds, but in the context of computer-based data systems it was felt that broad agreement on the intellectual content of geological files was basic to the operation of a decentralized Canadian System.

GEOSCIENCE DATA

CURRENT STATUS

The promotion and coordination of practical efforts to develop the concepts advocated by the National Advisory Committee report have been undertaken by the Canadian Centre for Geoscience Data since 1970 (Sharp, 1972). Its major activities during this period fall into three categories:

1. Inventory of data sources. Principally, continued development of the *Canadian Index to Geoscience Data*, with emphasis on various unpublished government reports (McGee, 1972), and more recently the identification and indexing of computer-based files (Gunn and White, in press; Hubaux, 1972).

2. Development of standards for geoscience data. Activity has been limited mainly to mineral and fuel deposits data, and to geological field data. For mineral deposits data, recommended standards are now being published (Kelly, 1972; Longe *et al.*, in press), but in the case of field data, emphasis has been placed on methods and techniques for developing successful computer-based systems for individual projects.

3. Coordination of data activities. The Canadian Centre for Geoscience Data has served as a focal point for inquiries, promoted contacts between workers, published bibliographies and other reports, and encouraged cooperation between government, universities and industry. The Centre also serves as secretariat for COGEODATA, the International Union of Geological Science's data committee. Research and development carried out under various auspicies, including its own, has been monitored and coordinated where possible.

Unfortunately, preoccupation with these day-to-day activities during the past three years has prevented the Centre from keeping in tune with an up-to-date conceptual framework. Technology has advanced so rapidly and in so many directions, that some of the basic assumptions underlying the Centre's programme of activities have become obsolete. The fact that there has been no general review of the Canadian System concept (Figure 1) since 1967 has, I believe, worked against the Centre's basic purpose of promoting the application of modern information technology in the most effective way possible. The Centre's Advisory Board will become more active in this area, but in the meantime I wish to present my assessment of the Canadian System concept as it exists today (Figure 2).

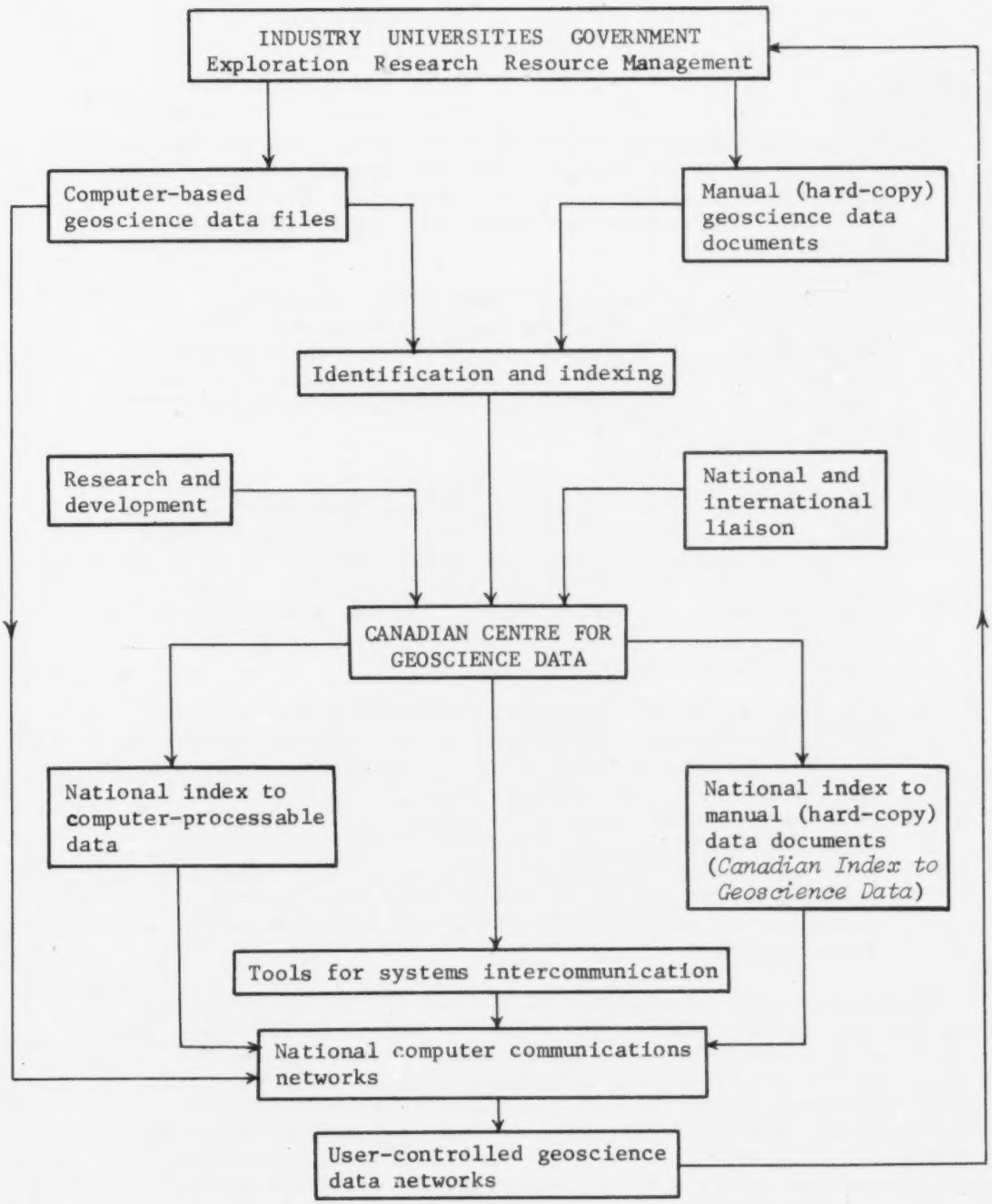

INDUSTRY UNIVERSITIES GOVERNMENT
Exploration Research Resource Management
Computer-based geoscience data files
Manual (hard-copy) geoscience data documents
Identification and indexing
Research and development
National and international liaison
CANADIAN CENTRE FOR GEOSCIENCE DATA
National index to computer-processable data
National index to manual (hard-copy) data documents (Canadian Index to Geoscience Data)
Tools for systems intercommunication
National computer communications networks
User-controlled geoscience data networks

Figure 2. Present concept of Canadian System for Geoscience Data as seen by the writer.

According to this conceptual framework, which I emphasize is not necessarily "official" doctrine, the Centre is responsible for three main products. One of these, called "Tools for systems intercommunication", a phrase used by UNISIST (H. Brown, 1971, p. 41), refers to the wide range of methods, technologies, thesauri, guides, and to a certain extent, standards, by which two or more data files may be used for a common purpose. This function replaces the single element "standards", perceived by the 1967 *ad hoc* committee as the only key to developing a national network of independant files. During the past six years, however, we have witnessed the development of many new tools, such as generalized data management systems; computers have become even more powerful, and programming and systems skills have eliminated many elements of technical incompatibility which were at one time considered insurmountable. Increasingly, we are able to accept files "as they are", and nevertheless use them in concert. An outstanding example familiar to many is the CAN/SDI service of the National Science Library (J. E. Brown, 1972). Thus, there is an increasing tendancy towards the view, shared by this writer, that standards for geoscience data, at least on the scale suggested in 1967, are no longer a prerequisite for national network development.

The two other main products of the Centre are indexes to sources of data on Canadian geoscience (Figure 2). I have concluded that two separate indexes are required because of the differences in indexing approach required for hard-copy documents and computer-processable files, respectively. Moreover, the latter tend to be dynamic in content and form, thus requiring frequent review and re-indexing.

The blend of these basic ingredients - indexes to data and tools for systems intercommunication - when combined with privately controlled data files operating within our increasingly sophisticated national computer communications networks, will, I predict, result in achievement of the goal sought by the National Advisory Committee in 1964 - a national, owner-controlled network of computer-based files for the use of all who may require geoscience data.

EVOLUTION OF CONCEPTS

The notion that modern information technology could be applied to assist in the national communication of geoscience observations and measurements was first advanced in 1964. A succession of individuals, committees, and organizations has contributed to the development of concepts by which this could be achieved. The essential problem they have faced is: given the objective of a decentralized, owner-controlled national data network, which components of the total system require a national consensus and national application? The answers have evolved

with time in response to advancing technology and have taken the following form:

Year	Key National Components	Technology	Reference
1965	Standard codes, formats, classifications, and descriptive terms, including geographic location. Standard reference numbering system.	80-column punched cards; card-sorting equipment.	Robinson (1965)
1967	Minimum standards for data content; standards for reference numbering, geographic location, and coding (if required). Permanent secretariat.	Computer systems; tape and disc storage.	Brisbin and Ediger (1967)
1973	File identification and documentation; tools for systems intercommunication (e.g. thesauri, tables of values, indexes, software); selected standards, esp. geographic location. Canadian Centre for Geoscience Data.	Computer systems; generalized data-base management systems; national computer communications networks.	This paper

The trend in this evolution has been away from the need for rigid standards for data content, notation and structure, and towards greater flexibility for data recording. Concomitantly, however, the same technology which is reducing the need for rigidity permits far greater practical use of data files among a wider audience. Such wide usage will, in turn, promote and bring about higher degrees of standardization among geoscientists. Thus, the application of information technology will, as a consequence, lead to greater standardization among geoscientists in response to legitimate scientific needs.

CONCLUSION

While it is generally appreciated that technology will continue to have a major effect on the evolution of national programmes for scientific and technical information, as illustrated by the Canadian System for Geoscience Data, the need for coordinating day-to-day activity with an up-to-date conceptual framework or model may be overlooked, and lead to serious delay and inefficiency. In practical terms, this danger can be avoided by the establishment of knowledgeable advisory and policy boards and by the maintenance of close working relationships between these bodies and practitioners of the art.

REFERENCES

BRISBIN, W.C. and Ediger, N.M., *Editors*, 1967, "A National System for Storage and Retrieval of Geological Data in Canada", National Adv. Comm. Research Geol. Sciences (available from Geol. Survey Canada), 175 p.

BROWN, H., 1971, "UNISIST: Synopsis of the Feasibility Study on a World Science Information System", Unesco, Paris 1971, 92 p.

BROWN, J.E., 1972, "The Canadian National Scientific and Technical Information (STI) System; a Progress Report", National Fed. Sci. Abstracting Services, Rept. 4, May 1972, 26 p.

GUNN, K.L. and WHITE, J.P.W., *in press*, "MINDEX: An Index to Mineral and Fuel Deposits Data Files in Canada", Can. Centre Geoscience Data, Geol. Survey Canada Paper 73-16.

HUBAUX, A., 1972, "Geological Data Files: Survey of International Activity", CODATA Bull. 8, 30 p.

KELLY, A.M., 1972, "Recommended Standards for Recording the Location of Mineral Deposits", Can. Centre Geoscience Data, Geol. Survey Canada Paper 72-9, 8 p.

LONGE, R.V. *et al*, *in press*, "Recommended Standards for Data on Mineral and Fuel Deposits in Canada (tentative title)", Can. Centre Geoscience Data, Geol. Survey Canada Paper.

McGEE, B.A., 1972, "New Key to Mineral Information - Canadian Index to Geoscience Data", Can. Mining Jour., v. 93, no. 4, p. 43-47.

ROBINSON, S.C., 1965, "Interim Report on Possible Application of Data Processing Techniques to Storage and Retrieval of Geological Data in Canada", unpublished rept., Geol. Survey Canada, Ottawa, January 1965, 79 p.

_____, 1966, "Interim Report of the Committee on Storage and Retrieval of Geological Data in Canada", National Adv. Comm. Research Geol. Sciences, Geol. Survey Canada Paper 66-43, 98 p.

SHARP, D.A., 1972, "The Canadian System for Geoscience Data", Proceedings 24th International Geol. Congress, sect. 16, p. 196-203.

THE POTENTIAL OF REGIONAL GOVERNMENT FOR THE RATIONALIZATION OF LOCAL INFORMATION SYSTEMS
(LE POTENTIEL OFFERT PAR UN GOUVERNEMENT REGIONAL DANS LA RATIONALISATION DES SYSTEMES D'INFORMATIONS LOCAUX)

J. Nicholas Brown, Senior Consultant,
P. S. Ross & Partners,
Toronto, Ontario M5H 3C8

ABSTRACT

The quality of local government information systems varies widely; the few good systems standing well above the milieu. The development of Regional Government in Ontario poses opportunities to improve this situation by offering a vehicle with the necessary economies of scale without loss of local autonomy. The potential role of the Region in information handling is examined and current experience directed towards realizing this potential is discussed. It has been found that the information required to manage the financial and administrative affairs of government can be met in a highly cost effective manner. The methodologies used to integrate existing routine applications and new applications making use of current information systems technology are described and these techniques are illustrated by examples for current situations. (La qualité des systemes d' information des gouvernements regionaux varie largement, et les quelques bons systemes ressortent aisement de la moyenne. Le developpement des gouvernement regionaux en Ontario crée des opportunites d'ameliorer cette situation en offrant un moyen ayant les economies d'échelles necessaires sans qu'il y ait perte d'autonomie locale. Le rôle potential de la region en ce qui a trait à l'administration de son information a eté examine et nous avons

discuté de quelle façon, l'éxperience presente tent à prouver ce potentiel. Il est prouvé que l'information requise, afin de gerer les aspects financiers et administratifs du gouvernement, peut être recueillie de façon trés effiçace quant au cout. Les methodes utilisées afin d'integer ces systemes existants et les nouvelles methodes utilisant la technologie la plus recente quant aux systems d'information, sont décrites et ces techniques sont illustrées par des exemples de situations courantes.)

POTENTIAL OF REGIONAL INFORMATION SYSTEMS

INTRODUCTION

Local government is a peculiar enigma: the focus of community participation, the seat of local autonomy, the bulwark against the encroachments of higher levels of government, and, at the same time, plaintiff for more funds to support locally administered projects.

The firm of P. S. Ross & Partners has had a continuing involvement with both large and small local governmental bodies. We are developing, in co-operation with the City of Toronto and the Province of Ontario, an extensive financial management and simulation system called PROMUS. We have served the Regions of Ottawa-Carlton and York and have assisted many cities and towns across Canada.

The complexities of the information requirements of local governments within metropolitan areas have been discussed at much lenght in other forums. I should like to concentrate upon the potential future development of local governmental information systems outside major metropolitan areas, and to identify the leadership required from senior levels of government if such projects are to be fully effective.

The Government of Ontario is establishing Regional Governments throughout southern areas of the Province. The principal differences between the regions and the former counties is that the constituent municipalities are an integral part of regional government. The new governmental functions are more extensive than those of the counties; regions are usually responsible for:

1. Police protection;
2. Road construction and maintenance (except in area municipalities);
3. Land use and transportation planning;
4. Water and sewage construction (except in area municipalities);
5. Health, Welfare and Social Services.

The constituent municipalities continue to administer:

1. Fire protection;
2. Local roads;

3. Internal water and sewage.

Still other agencies provide services for hydro electricity and education.

The regional services are financed in two ways:

1. By a levy from the constituent municipalities;
2. By Provincial grants for services rendered (i.e. roads, welfare).

This fragmentation of services may not be conducive to the most efficient use of information, but it is a standard division of responsibilities based on the premise that region-wide functions are the responsibility of the regional government and that local functions are the responsibility of the local government. Despite the seeming disadvantages of this division of responsibilities, the regional level of government turns out to be an economic and viable information processing unit. Potentially, regions are the nodes in a network of provincial information flows.

P. S. Ross & Partners has just concluded a study for The Regional Municipality of York, which is situated due north of Metropolitan Toronto. The total information requirements of the Region were analyzed and the potential for developing a comprehensive local information system explored. I should like to draw upon this experience as well as our recent experience at the Town of Mississauga to illustrate the potential for senior government leadership in guiding the development of local information systems. First, the general requirements for information are summarized, then a data structure well suited to meeting these requirements is described. Because of the potential cost of developing the necessary application programs, the contracting approach employed on behalf of The Regional Municipality of York is discussed. The precautions taken to overcome the inherent risks of this approach are explained.

A DEFINITION OF REQUIREMENTS

The information requirements of regional government centre around financial control, however, to manage a region effectively, much more than financial information is required. It is our experience that an information system, designed with an integrated

financial structure in mind, can provide as effective a mechanism for handling non-financial information as any other system while ensuring that the financial integrity of the system is maintained and is available for audit. Therefore, we have adopted an approach to system design which is illustrated in a simplified fashion in Figure I.

The heart of the financial system is, of course, the General Ledger. In general, each department has both revenue and expense accounts associated with the service it provides. Departments providing complex services require complex sub-ledgers to maintain control of these services. We found that the Engineering Department required a sub-ledger cpability to control both ongoing maintenance and new construction and to control the révenue entitlement earned under provincial cost sharing arrangements. The Health and Social Services Department required a detailed sub-ledger to handle the welfare payroll and the revenue entitlements generated from welfare payments.

From the above examples, it is apparent that the sub-ledgers contain much more than merely debits and credits; they are the master files for standard municipal applications.

Some applications have no financial component; for instance, Police have voluminous statistics on the occurance and dispositions of various criminal occurances which must be sorted, stored and reported. In The Regional Municipality of York there will be an opportunity to interface a digital communications network with the routine data processing functions thus greatly simplifying the analysis of the time spent by the various police constables in the performance of their duties.

A summary of the major Regional data processing applications by functional area is given in Figure II.

During our work at The Regional Municipality of York, it became apparent that any system to serve the Region would also have to have the capacity to serve the constituent municipalities. Furthermore, while there is some duplication of requirements, the main volume of data processing for the municipalities is associated with tax collection and utility billing. These applications, being oriented towards the land within the municipalities, complement the Region's requirement for engineering data for this land. That is to say, when

FIGURE I

SIMPLIFIED SYSTEM CONCEPT

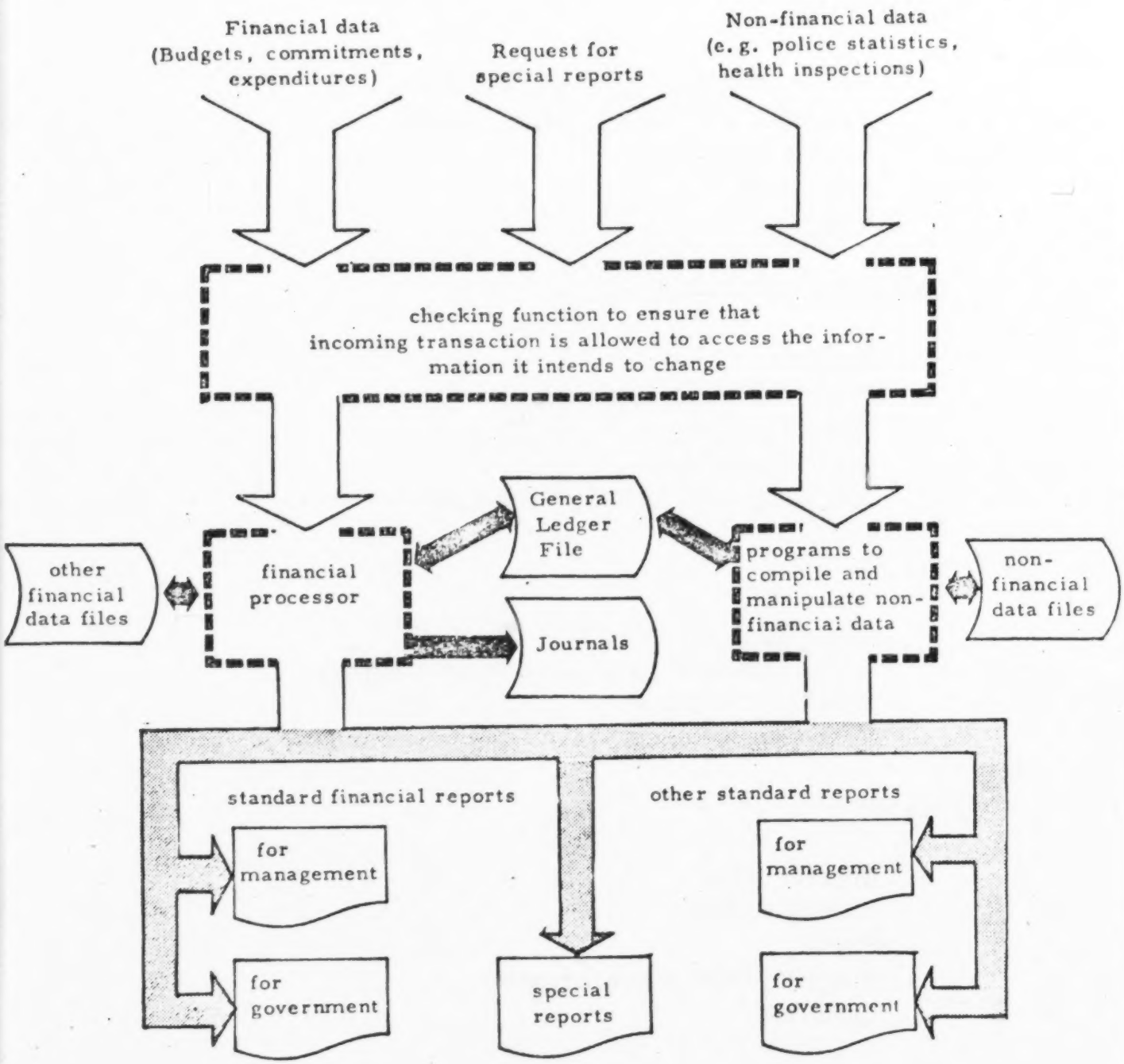


FIGURE II

AREAS OF POTENTIAL DATA PROCESSING APPLICATION WITHIN REGIONAL ORGANIZATION

ENGINEERING

- Design, Project Supervison and Control
- Maintenance for Roads, Water, Sewage
- Vehicle Maintenance
- Traffic Flow
- Property

HEALTH AND SOCIAL SERVICES

- Welfare
- Public Health Nursing
- Homes for Aged
- Health Inspection
- Special Sections (dental, daycare, etc.)

POLICE

- Crime Occurrence Records
- Traffic Summons Control and Accident Analysis
- Vehicle Maintenance
- Manpower Utilization

TREASURY

- General Ledger
- Subsidy Control
- Payroll
- Budgeting (with encumberances)
- Cash Flow

PLANNING

- Land Use

both Regional and Area requirements are viewed as a whole, there is a need for an integrated property-oriented data base. P. S. Ross has already explored the development of such a data base for the Town of Mississauga. The experience gained in Mississauga is reflected in the information structure described in the next section.

A summary of the data processing applications applicable to the municipalities is given in Figure III.

POSSIBLE INFORMATION STRUCTURE

Let us now look ahead to what type of structure would be best suited to meeting these information requirements. Ideally, the total Regional Information System would be organized as a functional data base with some elements of data being easily subdivided by municipality while other elements would contain only information pertinent to one municipality. Some of the components of this data base have been identified as follows.

General Ledger

Each municipality would have its own unique general ledger, however, the structure of each of these ledgers would be standardized. The revenue accounts would provide information on revenue entitlements earned as well as budgeted revenues and revenues actually received. In a similar manner, the expense accounts would support encumbrances as well as budgetary information and actual expenditures.

Special Ledgers

Functions such as engineering would require special ledgers to assist in the control of projects for both new construction and for maintenance as already described. If a region and several of its municipalities required engineering application, each user would have his own ledger, but all ledgers would have a similar structure.

Payrolls

The region and constituent municipalities would share an employee data base which provided personnel information and payroll records to each user. Integrating the processing of such functions while maintaining the ability to provide separate accounting reports

FIGURE III

AREAS OF POTENTIAL DATA PROCESSING APPLICATION OUTSIDE REGIONAL ORGANIZATION

AREA MUNICIPALITIES

- Water Billing
- Tax Billing
- Assessment Roll - Planning
- Any functions supported for Region (Slide 3)

ELECTRIC POWER DISTRIBUTION

- Billing, Accounts Receivable, Collections
- Vehicle Maintenance
- Manpower Utilization
- Work Order Control
- Payroll
- General Ledger

COUNTY HOSPITALS

- Payroll
- Inventory Control
- Capital Equipment Control

helps to realize the economies of scale available from the joint use of the data processing facility.

Resident Name and Address File

To simplify the process of maintaining the current name and address lists for residents of the region, such lists would be maintained in only one file and linked to other files by appropriate data base techniques.

Welfare Rolls

A simple file on all welfare recipients in the region could be maintained and would serve as a source of control for welfare payments and for other social service activities.

Physical Characteristics of Land

Such a file would describe location of every property, both taxable and exempt, within the region using the standard geocoding technique, and would provide such technical details as required by planning functions at municipal and regional levels, by engineering functions and by health functions.

Assessment Data

This file would be closely related to the property file and would provide all necessary information concerning assessed value of the land, the ownership of the land and the taxes due on the land. This data would allow the planners to view an integrated socio-economic picture of the region.

Data Reporting

The programs accessing such a data structure would facilitate preparation of reports not only to serve the business of municipal and regional administration but also to satisfy provincial requirements and to prepare entitlement claims for provincial funds under the various programs being sponsored by the province. The collection of the data into an integrated data base would in no way reduce the control of the respective municipal treasurers over their own funds. Transfers of funds between municipality and region and between region and the province would remain under the control of the appropriate treasurers. The proper organization of this data would simplify the

production of necessary managerial reports and allow the most effective use for the various civic resources such as health nurses, police, fire.

FROM HERE TO THERE

Since regional organization are new, they are usually unencumbered by existing information system structure. If a region can effectively resist the pressure to adopt a system similar to the one used in the county or to one used in one of the major constituent municipalities, it can develop an information system to suit its own requirements. When The Regional Municipality of York was first constituted, it built manual procedures around two antiquated accounting machines previously used by the county and then set about determining what its total requirements really were. The municipalities, which had a far longer history of civic government, were at various stages of sophistication. The pressure of tax billing and cash receipts had generally forced all municipalities to make some use of computer service bureaux in the handling of tax bills. The most sophisticated of these municipalities were finding that the local service bureau could not meet their processing requirements within their budgets and were actively considering acquiring their own computer.

Even without considering the requirements of the constituent municipalities, the potential systems design and programming tasks before the Region represented a major undertaking. Faced by the prohibitive costs of such an undertaking, we quickly found an alternative. None of the requirements that had been identified for either region or municipalities were unique and we reasoned that they could be met by existing programs. Our task therefore became one of defining requirements and finding the best source of supply of a total service, hardware and software.

Definition of Requirements

We set out to define our requirements so as to allow maximum latitude on the part of the supplier while at the same time maintaining sufficient exactness to ensure that there could be only one interpretation of the requirements. We quantified the volumes of inputs and frequency of outputs and identified the various elements of data that would be required in each. To ensure that the project would be carried out on schedule and within cost, we specified that a single contractor must take total responsibility for the whole project including

the supply of both computing services and software. No contraints were applied as to whether a utility or in-house computer would be used, however, it was specified that printing and data input had to be done on the premises of the Region.

Problems and Solutions

It was recognized that the extensive use of outside programming could result in the Region being left with the programs they did not understand. We elected to overcome this problem by insisting that the contractor use two members of Regional staff on his development team during implementation. To ensure that the variety of independent applications packages which the vendors were likely to propose would in fact be successfully meshed into a cohesive system, a number of constraints on software were laid out such that data names were to be common throughout the system and that input/output was to be performed by common modules. Not only would this meet the requirement of cohesion but also it would force the supplier to review his applications packages and allow the Region's staff to gain first-hand familiarity with them.

Tendering

Having defined the requirements in such a way as to make the maximum number of potential suppliers eligible and having set the conditions sufficiently stringent to protect the Region from any potentially inadequate bids, we invited a number of potential suppliers to tender. These suppliers included both hardware manufacturers and utilities. About one-third of the invitees responded with proposals. Of those proposals, several were acceptable in every regard. We ranked those proposals that were completely acceptable according to a number of predetermined criteria such as software capabilities, hardware capabilities and price.

CONCLUSIONS

I have demonstrated on the basis of the experience of P. S. Ross & Partners that the concept of regional government offers encouraging potential for uniting the information requirements of a number of local government bodies into a single body, and for realizing, not merely the economies of scale in computing, but also the economies of scale

of an integrated data base. While there is no doubt that the potential for greatly enhanced local government lies within the concept of regional government, it remains to be seen whether or not the spice of local politics overcomes the subtleties of fully integrated co-operative data processing on a regional basis.

LE SYSTEME CANADIEN DE RENSEIGNEMENTS SUR LA CONSTRUCTION (A CONSTRUCTION INFORMATION SYSTEM FOR CANADA)

Onil Dupuis
Société Cogito Limitée
50 Place Crémazie O., Suite 800, Montréal 351

ANALYSE

Le Système Canadien de Renseignements sur la Construction (SCRC) a pour but d'augmenter la productivité de l'industrie de la construction en améliorant la diffusion de l'information. Le repérage de l'information s'effectue par ordinateur, mais l'information est conservée sur microfiches. Il comprend quatre banques de données distinctes: produits et matériaux, techniques et méthodes, statistiques, codes et normes. (The main objective of the Construction Information System for Canada (CISC) is to increase productivity in the construction industry by providing better information. Information is searched by computer but information itself is on micro-fiches. Four different data bank are part of the system: products and materials, technics and methods, statistics, codes and standards).

INTRODUCTION

Malgré sa croissance rapide et la part imposante du produit national brut que représente l'industrie de la construction, il est devenu presqu'un lieu commun de lui reprocher son inefficacité, son extrême fragmentation, le coût exagéré de ses opérations et le caractère artisanal de ses méthodes. Par ailleurs, la croissance exponentielle de la masse d'information produite ou consommée chaque année par cette industrie n'a guère eu d'autres effets que celui d'augmenter la confusion et d'accentuer son caractère de fragmentation.

Le système Canadien de Renseignements sur la Construction (SCRC) est un système national d'information dont le but est d'augmenter la productivité de l'industrie de la construction en organisant de façon plus cohérente et plus efficace la collecte, le traitement et la diffusion des informations.

Après un bref historique de l'origine et du développement du SCRC nous décrirons dans ses grandes lignes l'organisation et le fonctionnement de ce nouveau système d'information. Enfin, bien qu'il soit trop tôt pour en faire une évaluation véritable, nous nous proposons d'en commenter les aspects qui nous semblent les plus intéressants ou, au contraire, qui présentent certaines faiblesses.

SCRC

HISTORIQUE

En 1967, le ministère de l'industrie et du commerce, dans le cadre de son programme BEAM, retenait les services de la firme Stevenson and Kellogg Limited afin d'étudier les méthodes de diffusion, d'emmaganisage et de recherches des renseignements en usage dans l'industrie de la construction. Le rapport de cette enquête révélait que 3% de la valeur totale de tous les travaux de construction d'une année, soit près de 300 millions de dollars, étaient annuellement consacrés à la production, la diffusion, l'emmagasinage et au repérage de l'information dans l'industrie de la construction. Uniquement sur les produits et les matériaux de construction, il se diffuse chaque année entre 20 et 30 millions de pages de documentation, et le rythme de croissance de cette documentation est de 8%... Malgré le fait que 65% de cette documentation soient jetée dès réception, ou après une première lecture, l'industrie de la construction canadienne n'en conserve pas moins dans ses locaux un total de 400 millions de pages de documentation sur les produits et les techniques de construction.

A la suite de ce rapport, la firme Demers, Gordon and Baby était chargé en 1969, de la conception et de la mise sur pied d'un système pilote d'information. Celui-ci fut mis à l'essai par 42 firmes de la région comprise entre Toronto, Ottawa, Montréal et Québec. Pendant ce temps, divers sous-contrats avaient été octroyés afin de développer les outils documentaires nécessaires (thésaurus, paramètre de produits, etc...), et de commencer à compiler les informations destinées à constituer le noyau initial des différentes banques de données.

En 1972,une version corrigée du SCRC était présentée aux membres de l'industrie de la construction, ses futurs usagers, au cours d'une conférence nationale qui eut lieu à Ottawa. Un peu plus tard, la Société Canadienne d'Information sur la Construction, une société à but non-lucratif, était créée afin d'assurer l'exploitation et l'expansion du système. Un deuxième et dernier test du système est prévu pour septembre prochain après quoi, vers la fin de novembre 1973 ou vers le début de janvier 1974, selon l'importance des modifications à lui apporter, le SCRC sera accessible au public.

ORGANISATION ET FONCTIONNEMENT

La conception du SCRC a été réalisé en fonction de quatre principaux critères. D'abord, l'information doit être facilement accessible à tous sous une forme immédiatement utilisable. Deuxièmement, le système doit être viable techniquement, i.e. conçu de telle sorte qu'il puisse profiter des possibilités de la technologie actuelle mais être suffisamment flexible pour pouvoir s'adapter au progrès futur de la

technologie. Troisièmement, le système doit être économiquement rentable. Enfin, il doit être en mesure d'offrir ses services indifféremment en français ou en anglais.

Le SCRC s'adresse à tous les membres de l'industrie du bâtiment depuis le manufacturier de produits et de matériaux de construction jusqu'à l'entrepreneur, en passant par l'architecte et l'ingénieur.

Fondamentalement, le système se compose de quatre banques de données, ou plutôt, contient quatre types différents d'information qui correspondent à quatre phases successives du plan global d'implantation du système. La première est une banque de données sur les produits et les matériaux. Celle-ci, bien qu'en retard de plusieurs mois, sera probablement terminée de compiler et disponible vers le mois de septembre prochain lors du test final du système. Dès novembre, les usagers pourront y avoir accès librement. Selon les prévisions actuelles, elle contiendra des données sur environ 800 produits architecturaux (enveloppe du bâtiment) et un certain nombre de produits et de matériaux électriques ou mécaniques. Ces deux derniers seront compilés d'une façon plus exhaustive au cours de 1974.

La deuxième banque contient des données sur les méthodes et les techniques de construction. En plus de la référence bibliographique, cette banque inclut une analyse de chaque document indexé. Bien que 3,000 documents aient déjà été indexés et incorporés à cette banque, le plan global d'implantation du système ne prévoit pas la rendre accessible avant la fin de 1974.

Les troisième et quatrième banques contiendront respectivement des informations commerciales et statistiques pour la première, et les normes, les codes, les lois pour la deuxième. Ces dernières ne seront mises en marché qu'au cours de l'année 1975.

La nécessité de rendre le système accessible à tous les usagers, indépendamment de leur situation géographique ou de la taille de leur firme, le coût élevé des communications ainsi que du temps ordinateur ont déterminé le choix du type d'équipement. Le SCRC est un système semi-automatisé.

Toutes les informations ou les données sont sur des microfiches distribuées à tous les usagers. Seul le repérage de l'information s'effectue par ordinateur, à l'aide d'un code numérique, au moyen d'un terminal situé chez l'usager. Les résultats de la recherche donnent à l'usager l'adresse alphanumérique de la microfiche sur laquelle il trouvera l'information désirée. Deux index imprimés (index des produits, index des fournisseurs), et deux répertoires (répertoire des marques commerciales, répertoire des paramètres des produits) complètent le système.

En n'automatisant que partiellement le système, il a été possible de répondre à deux objectifs majeurs du SCRC: rendre l'information accessible au plus grand nombre possible d'usagers et maintenir le service à son coût le plus bas. Il en coûtera entre $50 et $200 par mois pour utiliser le système. Ce prix inclut le coût de location du terminal et de la visionneuse, aussi bien que le coût des communications et du service.

Le financement du SCIC sera assuré grâce à la vente du SCRC aux membres de l'industrie de la construction et grâce aux cotisations des manufacturiers dont les produits seront inclus dans la banque de données. Le SCIC estime être en mesure de s'autofinancer dès 1975 ou 1976.

OBSERVATIONS CRITIQUES

Cette troisième partie n'a d'autres buts que de faire une critique rapide de quelques aspects du SCRC. Il est trop tôt pour en faire une évaluation véritable puisque le système n'est pas encore opérationnel. De plus, comme nous sommes limités dans le temps qui nous est dévolu, nous avons choisi de ne commenter que brièvement la banque de données sur les produits et les matériaux (sa plus grande faiblesse réside au niveau de sa mise à jour continuelle; il est difficile et coûteux de remplacer des index imprimés...) pour nous attarder davantage à la banque de données sur les méthodes et les techniques de construction (à laquelle nous reprochons surtout d'avoir été conçue de la même façon que la banque de données sur les produits et les matériaux).

La banque de données sur les produits et les matériaux de construction est compilée uniquement à partir des informations recueillies auprès des manufacturiers. Le SCRC n'a aucun contrôle sur la qualité, l'exactitude ou l'exhaustivité des renseignements qui y sont stockés. Son rôle se limite à normaliser la forme sous laquelle se présente l'information. Il s'agit cependant ici d'un grand pas en avant puisque pour la première fois, l'architecte, l'ingénieur ou l'entrepreneur est en mesure de comparer deux produits similaires. Avec la rapidité avec laquelle l'information est accessible, il s'agit ici du principal avantage de cette banque de données et c'est dans ce sens qu'elle est beaucoup plus qu'un Sweet's Catalog automatisé.

Le SCRC distribue à tous ses usagers les informations sur les produits au moyen de microfiches. Par contre, en ce qui concerne la banque de données sur les techniques et les méthodes de construction, seule l'analyse du document est disponible sur microfiche. L'usager doit se procurer lui-même le ou les documents qui l'intéressent: ce qui implique des coûts et des délais supplémentaires que celui-ci ne veut ou ne peut pas se permettre et qui diminue d'autant pour lui la rentabilité d'un tel système d'information.

SCRC

L'objectif premier du SCRC étant d'améliorer la productivité de l'industrie de la construction en mettant à sa disposition l'information dont il a besoin au moment où il en a besoin, il nous semble que cet objectif serait plus facilement atteint si l'usager avait la possibilité d'emprunter d'une bibliothèque dépositaire ou, mieux, d'obtenir d'un centre de distribution une microfiche de ce document.

La nécessité de ce centre de distribution nous paraît s'imposer d'autant plus que l'analyse d'un document (du moins dans un domaine tel que l'industrie de la construction) permet uniquement d'évaluer la pertinence d'un document par rapport à une question mais ne donne pas d'information comme tel. De plus, mais ce problème aura certainement trouvé une solution avant que la banque ne devienne accessible au public, la qualité des analyses nous a paru fluctuer de façon considérable.

Au niveau des outils documentaires, le thésaurus des termes de la construction tel que développé par Hanscombe, Roy and Associates et dont la version française est parue en 1972, convient assez bien à l'indexation des données sur les produits et les matériaux, en dépit du fait qu'il ne s'agisse pas d'un vocabulaire contrôlé dynamique puisqu'aucun mécanisme de mise à jour continuelle n'a été jusqu'à maintenant prévu. Par contre, son utilité est plus que discutable en ce qui concerne l'indexation des données sur les méthodes et les techniques de construction puisqu'on y trouve que 20% des descripteurs utilisés. Le thésaurus employé pour cette banque de données a été celui développé par Industrialisation Forum dont M.C.H. Davidson vous a entretenu hier.

En plus du problème de la traduction (en cours de réalisation) et de la réconciliation des deux thésaurus (ce qui ne présente pas de difficulté majeure puisque 20% des termes sont communs, 70% sont mutuellement exclusifs et 10% seulement sont synonymes ou quasi-synonymes), l'utilisation de ce deuxième vocabulaire contrôlé a soulevé un problème nouveau. La complexité et la sophistication de sa structure permettent un haut degré de raffinement aux deux étapes de l'indexation et du répertoriage, mais nécessite l'intervention d'un intermédiaire entre l'usager et la banque. Des tests préliminaires ont, en effet, mis en évidence la difficulté rencontrée par l'usager à obtenir un niveau de pertinence satisfaisant des informations. Cette difficulté s'aggrave encore du fait des différences considérables de formation et de structure de pensée entre les différents usagers.

L'introduction d'un tel intermédiaire entre l'usager et la banque de données suppose bien sûr des coûts supplémentaires. Ceux-ci seraient cependant facilement absorbés par un taux de rentabilité accrue, d'autant plus qu'il serait ainsi possible d'utiliser les bandes développées par d'autres systèmes d'information dans des domaines connexes telles celles

de lSl ou de Engineering Index. Le SCRC n'avait ni envisagé cette possibilité ni prévu de modalités de collaboration et d'échanges avec d'autres systèmes d'information nationaux similaires en voie de formation en Grande-Bretagne et en Scandinavie.

CONCLUSION

Les quelques remarques qui précèdent n'ont aucunement l'intention de faire une critique négative du SCRC dans son ensemble, mais, au contraire, de suggérer quelques modifications susceptibles d'accroître sa rentabilité. Nous sommes persuadés que l'industrie de la construction canadienne profitera d'une façon sensible de l'implantation du SCRC, même dans sa forme actuelle. Ce profit se traduira par une plus grande cohésion et une collaboration plus étroite entre ses membres, d'où une efficacité, une productivité et une rentabilité accrue.

BIBLIOGRAPHIE

Canada, Department of Industry, Trade and Commerce, 1971, Thésaurus of Canadian Construction Terminology. Preliminary edition, Ottawa, Department of Industry, Trade and Commerce.

Industrialisation Forum, 1972, Thésaurus of Building Science and Technology, Preliminary edition. Montréal, Industrialisation Forum.

"Information system officially unveiled", 1972, Architecture Canada, vol. 59, no. 4, p. 4.

National Conference on Construction Information, Proceedings, 1972, Ottawa, Department of Industry, Trade and Commerce, 141 pp.

Nicol, Frank, 1971, "Field trials of the BEAM Construction Information System", Architecture Canada, vol. 58, no. 5, pp. 6-7.

ANNEXE

APPENDIX

Autres articles soumis
Other contributed papers

1. ETHNOLOGIE, OEUVRES D'ART ET MÉTIERS DU QUÉBEC

André Barbeau, Emile Loranger
Centre de documentation
Bibliothèque
Université Laval, Québec

Fruit de trois ans de travail, le service d'information et de documentation sur l'ethnologie, les oeuvres d'art et les métiers du Québec (EOM) est issu de 3 fonds distincts: le fonds photographique, le fonds des artisans et l'ethnologie de l'Amérique française. Le fonds photographique est constitué d'environ 35,000 photos (3" x 5" et 8" x 10") et près de 25,000 grilles d'analyse portant sur les oeuvres d'art du Québec: peintures, églises, maisons, sculptures, etc. ... La nature même de cette documentation exigeait le choix d'un support d'information tel que le sélecteur optique. La grille d'analyse a été conçue en fonction des huit rubriques (référence, collection, auteur, époque, catégorie, sujet, identification, caractéristique (décor et matériau)) qui regroupent chacune un vocabulaire descriptif, libre et neutre. L'importance du vocabulaire varie selon la rubrique. Ce vocabulaire est mis en ordre alphabétique et chaque nouveau mot reçoit un numéro séquentiel qui deviendra la codification sur la grille d'analyse. Quelques problèmes se sont présentés au niveau de cette codification en particulier en ce qui concerne les champs. La conception de la grille d'analyse a dû tenir compte des limites physiques du sélecteur optique (Miracode). Le fonds photographique comprend 50 cassettes de microfilm, un lexique contenant 7 index et plus de 15,000 entrées lexicales qui permettent de consulter ce fonds à l'aide du sélecteur optique. Les usagers de ce service sont aidés dans leur recherche documentaire par un informatiste du Centre de documentation. Le développement de cette banque d'information se fait par les mises à jour du fonds photographique lui-même et par l'élaboration d'autres projets portant sur un aspect plus particulier de l'ethnologie, des oeuvres d'art et métiers du Québec.

2. CRÉATION D'UN RÉPERTOIRE ANALYTIQUE DE LANGUE FRANCAISE

Raymond Blais, Raynald Francoeur, André Gamache
Centre de documentation
Bibliothèque
Université Laval, Québec

Une nouvelle structure de répertoire analytique a été mis sur pieds au Centre de documentation de la Bibliothèque de l'Université Laval. Elle s'inspire d'une critique de la formule antérieure des principaux répertoires fabriqués au Centre de documentation: Index analytique, Index de l'actualité vue à travers le presse écrite, Répertoire des thèses de doctorat soutenues devant les universités de langue française... Elle répond à 4 critères précis: accès multiple à un même texte, réduction maximale des renvois, analyse approfondie des textes, réduction maximale des coûts. Elle a été confirmée par une étude de 11 index en sciences humaines. Le nouveau répertoire distribue l'information dans deux sections complémentaires: analytique et méthodique. Il est compilé par ordinateur grâce à la programmerie CARA. Cette nouvelle structure sert à la préparation de "RADAR", Répertoire analytique d'articles de (100) revues du Québec.

3. THE MAINTENANCE OF ON-LINE COMPUTER LIBRARIES

David A. Bonyon
I.P. Sharp Associates
320 Queen Street
Ottawa K1R 5A3

Computer technology and reliable communications have made possible time-shared networks over wide geographical areas. The creation, maintenance and use of shared on-line libraries within these systems points the way to the use of a network of computers to allow a user to access information in an on-line library remote from his own computer. Before one attempts to link computers for information reasons (they are already being linked for other purposes such as alternative processing capabilities) it would seem advisable to take a close look at the problems and successes of the on-line library in a time-shared network. This paper does this with reference to the APL Plus system of I. P. Sharp Associates Limited.

4. ÉTABLISSEMENT D'UN RÉPERTOIRE POUR CENTRE DE DOCUMENTATION À L'AIDE DU PROGRAMME DE IBM KWIC

Jean Bouthillette
Bibliothèque Générale
Université de Sherbrooke
Sherbrooke, Québec

5. AN INFORMATION SYSTEM TO IMPROVE PRODUCTIVITY AND EFFICIENCY IN INDUSTRY

John A. Dawson
Construction Division
Materials Branch
Department of Industry,
Trade and Commerce
Ottawa

6. VIBANQUE

Jean-Eudes Desgagnés, Srinivasan Venkatasubbu, Emile Loranger
Centre de documentation
Bibliothèque
Université Laval, Québec

VIBANQUE est une banque d'information spécialisée sur la mécanique des vibrations. L'organigramme de ce système comprend les étapes suivantes: acquisition, sélection, duplication, indexation, microcopie, saisie des informations. Cette banque d'information renferme près de 15,000 résumés analytiques avec un taux de croissance annuel d'environ 5,000. Le service VIBANQUE sert à alimenter la recherche à l'Université Laval dans le domaine de la mécanique des vibrations et est offert à l'ensemble des chercheurs canadiens ou étrangers. L'interrogation se fait à partir des éléments de dépistage suivants: mots du titre, année de publication, auteurs, descripteurs méthodiques.

7. CURRENT AWARENESS AT STELCO

Arthur D. Dick, R. D. Saunders
Steel Company of Canada Limited
Wilcox Street
Hamilton 23, Ontario

A computer-based information announcement system for technical and management personnel at Stelco has been operating for three years. Incoming literature is scanned and evaluated by technical personnel and submitted to the Research Library for coding and subsequent matching against personalized profiles of 220 individuals across the company. Weekly bulletins are prepared and resulting requests for the announced literature are filled from a master file.

8. WHAT IS S.C.T.I.E.?

Pierrette Dubuc
Centre d'Information sur l'Enfance
et l'Adolescence Inadaptées
Département de Psychiatrie
Hopital Sainte-Justine
3100 Ellendale
Montréal 251, Québec

The "Service de compilation et de traitement de l'information en éducation" has originated from an informal communication between a researcher and an information specialist. Collaboration has made it grow from a project report to an experiment actually going on at "CADRE" (1) and financed by the Department of Education of Quebec.

Inspired by the ERIC system, it has one original character: it is a cooperative system and implies the authors' participation. Based on the ERIC descriptive form, adaptable to the MARC II format, it is proposed actually to the Quebec researcher as a standard unit from which, with the help of the information specialist, a network can grow. A bilingual Thesaurus will facilitate the exchange with ERIC and other centers.

(1) Centre Animation - Développement - Recherche en Education, Montreal.

9. COMPUTER IDENTIFICATION OF LEVELS OF STYLE IN SCIENTIFIC AND TECHNICAL WRITING AND ITS IMPLICATION FOR IMPROVED ANALYSIS OF INFORMATION TRANSFER IN CANADA

Doreen M. Heaps, Wayne D. Ingram
Department of Computer Science
Sir George Williams University
Montreal, Quebec

A discussion is given of the transfer of scientific and technical information in Canada through publications that range from journals of learned societies to technical news items in daily newspapers. It is postulated that only particular sections of the publication range reach makers of public policy. It is proposed that efficient matching of users to appropriate data bases and improved transformation of scientific and technical information from one level of discourse to another would aid in the formation of wise science policy. Such matches and transformation may be aided by automatic means. An account is given of the automatic identification and representation of "levels of discourse" in samples of scientific and technical writing, including examples from certain typical Canadian publications. Conclusions from the analysis and implications both for future information networks and for public policy formation are given.

10. A GENERAL PURPOSE PROGRAM CAN ELIMINATE OPERATING SYSTEMS AND CONVENTIONAL PROGRAMMING LANGUAGES FOR INFORMATION SYSTEMS DESIGN AND DRASTICALLY CUT DEVELOPMENT COSTS

Bernard A. Hodson
School of Business Administration
Lakehead University
Postal Station P
Thunder Bay, Ontario

11. INFORMATICS UNLTD.: THE FIRST ANNUAL S.R. RANGANATHAN MEMORIAL INSURRECTION

Stephen Hollander
Health Protection Branch
Health and Welfare Canada
Box 3174
Station C
Ottawa K1Y 4J4

The reason that information theory today is so bastardized and faddish is that it lacks a humanistic discipline. Leaving aside the cookbooks which tell us "how I run my system good" and the pie-in-the-sky visions of poops whose meanderings are made none the less fatuous by the fact that they have, on rare occasions and by sheer accident, proven to be correct, the existing documents dealing with the theoretical aspects of information are either trivial or self-evident, and most are both.

12. AUTOMATIC INDEXING OF SEMINAR REPORTS: A TOOL FOR THE CLASSROOM AND THE CURRICULUM

William H. Hudgins
School of Library and Information Science
University of Western Ontario
London, Ontario N6A 3K7

The paper will describe briefly a keyword indexing system written in BASIC which is used to summarize student reports for seminars at the School, and will comment upon its use in course content development and evaluation. The system will have been in use for a full term by the time of the Conference.

13. L'INDEX DE L'ACTUALITÉ VUE À TRAVERS LA PRESSE ÉCRITE

André Khordoc, André Barbeau, Jean-Eudes Desgagnés
Centre de documentation
Bibliothèque
Université Laval, Québec

L'objectif de l'Index de l'actualité est de fournir aux canadiens un instrument pour dépister l'information portant sur l'actualité canadienne et internationale. Les sources auxquelles s'alimente l'Index sont "Le Devoir" et "La Presse", de Montréal, ainsi que "Le Soleil" de Québec. La sélection des articles à analyser se fait selon les critères suivants:- on retient les éditoriaux et les suppléments littéraires des trois quotidiens. - Pour le Devoir: on retient les nouvelles locales, nationales et internationales, les libres opinions, les bloc-notes, les collaborations spéciales et les cahiers spéciaux. - On rejette les fluctuations boursières et du marché, les événements sportifs, les annonces de spectacles, la rubrique: "Les tribunaux" et les lettres des lecteurs lorsqu'elles ne présentent pas un intérêt particulier. Le langage descripteur utilisé pour l'indexation est uniformisé périodiquement de manière à correspondre à la réalité actuelle de l'information. L'Index est préparé par ordinateur à l'aide de la programmerie EXACT dont l'une des étapes consiste à justifier le texte selon les règles de syllabation de la grammaire française. Les caractéristiques de cet Index sont: a) l'importance de son

contenu canadien, b) sa périodicité mensuelle et sa cumulation annuelle (signalisation rapide et pratique de l'actualité), c) la primauté dans son accès des concepts analytiques variables sur le schéma méthodique plus rigide. Au-delà de l'imprimé, les informations enregistrées sur support magnétique depuis 1966 constituent une banque d'information qui peut être interrogée par ordinateur.

14. INFORMATION RETRIEVAL IN MATHEMATICS
CANADA COULD TAKE THE LEAD

Kenneth O. May
Department of Mathematics
University of Toronto
Toronto 5, Canada

15. BUILDING & MAINTAINING A NATIONAL WATER RESOURCES DATA BASE FOR CANADA: A CO-OPERATIVE APPROACH

Marcel A. Mercier, Janice Heyworth
Environment Canada
Water Resources Data Systems
Inland Water Directorate
No. 8 Temporary Building
Ottawa

The establishment and maintenance of an on-line nationwide water resources data base which is distinctively Canadian in character depend upon the regular contributions of those who use it. Methods of locating and acquiring published and unpublished documents and current research information have been devised to ensure the data base is comprehensive and up-to-date. The information is coded, indexed, and abstracted for input to the system according to the requirements for efficient retrieval from a large on-line system. The need for compatibility between systems and the feasibility of exchange agreements of data banks both within and outside Canada influence the design criteria.

16. WHAT TO TEACH THEM IN GRADUATE SCHOOLS OF INFORMATION SCIENCE

S. D. Neill
School of Library and Information Science
The University of Western Ontario
London, Canada

Graduate education is different from undergraduate education. In undergraduate education students are taught knowledge about a subject, are taught to understand that knowledge (content), and are taught how to apply it. Graduate education teaches how to analyse the primary materials in the subject field, how to synthesize, and how to evaluate. In other words, graduate education is concerned with the means of finding answers, and not the answers alone. This paper is, therefore, an attempt to describe the 'discipline' of library and information science. Before one decides how to teach something, one needs to know what it is one wants to teach.

17. A GOVERNMENT DOCUMENTS DATA BASE - QUICK, CHEAP, AND CANADIAN

Merv Hrabi
Library Systems and
Data Processing
University of Guelph Library
Guelph, Ontario N1G 2W1

Ellen M. Pearson
Documentation Centre
University of Guelph Library
Guelph, Ontario N1G 2W1

Basic design decisions; how the system was implemented; present system operation, and problems; future possibilities: SDI, co-operative input and processing.

18. A MATHEMATICAL MODEL OF AN INFORMATION SYSTEM

F. W. Slingerland
Département de Génie mécanique
Université Laval, Québec

Srinivasan Venkatasubbu
Centre de documentation
Bibliothèque
Université Laval, Québec

The unit cost, coverage and cost-effectiveness of a specialized data bank in an exponentially growing literature field are analyzed, taking into account decay of interest with time past. Formulae are evolved which permit system assessment and policy guidance on: - balance of expenditure between current accumulation and retrospective build up; - initiation and termination of retrospective input. The application of the theory is illustrated for the project "Vibank", and information system on mechanical vibration.

19. A CONTRIBUTION TO THE DEVELOPMENT OF INFORMATION SERVICES IN CANADA

O. R. Standera
Information Systems, Library
The University of Calgary
Calgary, Alberta T2N 1N4

20. AUGUR RETRIEVAL SYSTEM

Jean-Pierre Theoret
Bell Canada/SGWU
620 Belmont Street
Montreal, P.Q.

"AUGUR" is a self contained system which sets up a data base from a sequential file of fixed length records, processes complex interrogations, and produces reports on qualified information.

The system is designed to process files incorporating relatively large records (25 to 10000 data items/record), where the number of records may range from 100 to 10000, and wherein any of the items may be used as a key. Updating at more than an item level should be infrequent.

These specifications enable the incorporation of a data structure free of pointers, and a very efficient, powerful retrieval procedure able to perform arithmetic as well as logical operatives, all with a minimal number of file accesses. It produces either a sequential file or printout of information, sorted by any number of keys. In addition, it allows default or defined synonyms.

Although the system is not applicable to all data bases, its simplicity and power makes it a valuable tool for specific applications.

21. INTERNATIONAL INFORMATION SYSTEMS IN ECONOMIC AND SOCIAL DEVELOPMENT: IDRC PARTICIPATION*

Kate Wild
International Development Research Centre
Ottawa, Canada

Information is necessary in the development process, from policy formation to project evaluation. To draw together all sources and users of development information, a co-operative network is required.

A "Macrothesaurus" or nucleus of development terminology has been compiled, with provisions for switching languages, and plans for switching to "micro" lists of more specific terms.

The IOB (Inter-Organization Board for Information Systems and Related Activities) was created in 1971 by the UN to establish a management information system for development activities. The IOB is working on the design of COMPASS, a development project information system utilizing the framework of the Macrothesaurus, and capable of utilizing input from, and serving the needs of, the UN Family, various aid-giving organizations, and as well policy-making, planning, and project managing agencies in the developing countries.

One aspect of the program of IDRC's Information Sciences Division is the support of these interorganizational activities.

22. MANAGEMENT INFORMATION SYSTEMS EDUCATION AT UNIVERSITY OF BRITISH COLUMBIA

H. J. Will
Faculty of Commerce &
Business Administration
University of British Columbia
Vancouver, British Columbia

*To be published elsewhere by author.

23. PREPARATION AND OPTIMIZATION OF SDI PROFILES FOR A DISPERSED COMMUNITY OF USERS

E. A. Williams
DSIS
Defence Research Board
125 Elgin Street
Ottawa K1A 0Z3

The preparation and servicing of profiles is a major part of the cost of an SDI service. Optimization of the profile takes several iterations, and this can be a lengthy process if it is geared to the normal SDI cycle. The process can be hastened and made more satisfying to the user if special SDI runs with short turnaround time can be provided during workshops held at the users' site. The approach used in the SOCRATES system is described. The advantages of profile entry from remote terminals as well as the flexibility of minicomputers in this application, are analyzed.